
CMS Physics Analysis Summary

Contact: cms-pog-conveners-pflowtau@cern.ch

2009/04/28

Particle–Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss}

The CMS Collaboration

Abstract

The versatility of the CMS apparatus is exploited in an attempt to identify and reconstruct individually each particle arising from LHC proton-proton collisions with a combination of the information from all sub-detectors. The resulting particle-flow event reconstruction leads to an improved expected performance for jets, taus and missing transverse energy.

Contents

1	Introduction	2
2	Fundamental elements	3
2.1	Iterative Tracking	3
2.2	Calorimeter Clustering	5
3	Link algorithm	5
4	Particle reconstruction and identification	6
4.1	Description of the algorithm	6
4.2	Cluster calibration	8
5	Performance of the Particle-Flow Event Reconstruction	9
5.1	Performance with Jets	10
5.2	Performance with Taus	18
5.3	Performance with Missing Transverse Energy	19
6	Conclusion	21

1 Introduction

The particle-flow event reconstruction aims at reconstructing and identifying all stable particles in the event, *i.e.*, electrons, muons, photons, charged hadrons and neutral hadrons, with a thorough combination of all CMS sub-detectors towards an optimal determination of their direction, energy and type. This list of individual particles is then used, as if it came from a Monte-Carlo event generator, to build jets (from which the quark and gluon energies and directions are inferred), to determine the missing transverse energy E_T^{miss} (which gives an estimate of the direction and energy of the neutrinos and other invisible particles), to reconstruct and identify taus from their decay products, to quantify charged lepton isolation with respect to other particles, to tag b jets, etc.

The CMS detector [1] appears to be almost ideally suited for this purpose. With its large silicon tracker immersed in a uniform axial magnetic field of 3.8 T provided by a superconducting solenoidal coil, charged-particle tracks can be reconstructed with large efficiency and adequately small fake rate down to a momentum transverse to the beam (p_T) of 150 MeV/c, for pseudo-rapidities as large as ± 2.6 .

Photons are reconstructed with an excellent energy resolution by an essentially hermetic electromagnetic calorimeter (ECAL) surrounding the tracker and made of over 75,000 $2\text{ cm} \times 2\text{ cm}$ PbWO_4 crystals, all the way to pseudo-rapidities of ± 3.0 . Together with the large magnetic field, the ECAL granularity is a key element in the feasibility of a particle-flow event reconstruction at CMS, as it generally allows photons to be separated from charged-particle energy deposits even in jets with a p_T of several hundreds of GeV/c. The granularity is even enhanced by an order of magnitude in the end-caps by the presence of a lead and silicon-strip pre-shower (PS) in front of the ECAL crystals.

Charged and neutral hadrons deposit their energy in the hadron calorimeter (HCAL) made of brass and scintillators, installed inside the coil and surrounding the ECAL, with a similar pseudo-rapidity coverage. The granularity of the HCAL is 25 times coarser than that of the ECAL, which would not allow charged and neutral hadrons to be spatially separated in jets with a transverse momentum much above 100 GeV/c. The hadron energy resolution in the combined ECAL–HCAL system is, however, of the order of 10% at 100 GeV. This resolution allows neutral hadrons to be detected as an energy excess on top of the energy deposited by the charged hadrons pointing to the same calorimeter cells. The charged hadrons are reconstructed with the superior angular and energy resolutions of the tracker. Particles with pseudo-rapidities between 3.0 and 5.0 are more coarsely measured with an additional forward calorimeter (HF), placed 11 m from the interaction point.

Electrons are reconstructed by a combination of a track and of several energy deposits in the ECAL, from the electron itself and from possible Bremsstrahlung photons radiated by the electron in the tracker material on its way to the ECAL. Muons are reconstructed and identified, in isolation as well as in jets, with very large efficiency and purity from a combination of the tracker and muon chamber information. Finally, the presence of neutrinos and other weakly-interacting particles can be detected by transverse missing energy (E_T^{miss}), defined as the modulus of the vector sum of the transverse momenta of all reconstructed particles.

The focus is put in this note on the reconstruction of charged hadrons, neutral hadrons and photons, which are the basic constituents of hadronic jets and of hadronic tau decays, and from which the E_T^{miss} is determined. The note organization follows closely that of the particle-flow event reconstruction algorithm itself. The reconstruction of its fundamental “elements”, the charged-particle tracks and the calorimetric clusters, is described in Section 2. These elements

are then topologically linked into “blocks”, as explained in Section 3. The core of the particle-flow algorithm, presented in Section 4, interprets the blocks in terms of particles. Finally, the expected performance of the particle-flow reconstruction in terms of jets, taus and E_T^{miss} are summarized in Section 5.

To ease the visualization of the detector dimensions and granularity, as well as the understanding of the basic principles of the particle-flow event reconstruction, an event display of a very simple hadronic jet with four particles (π^+ , π^- , π^0 , K_L^0) and a p_T of 65 GeV/c is shown in Fig. 1.

2 Fundamental elements

Most stable particles produced in proton-proton collisions have a rather low p_T , even in collisions involving a large momentum transfer. For example, in a quark or gluon jet with a total p_T of 500 GeV/c, the average p_T carried by the stable constituent particles is of the order of 10 GeV/c. This value reduces to a few GeV/c in jets with a total p_T below 100 GeV/c, typical of the decay chains of heavy exotic particles. To disentangle the production of these exotic particles from the dominating standard-model background processes, it is therefore essential to accurately reconstruct and identify as many of the final stable particles as possible, some of them with small p_T 's and energies.

As mentioned above, the particle reconstruction and identification is performed with a combination of the information from each CMS sub-detector, under the form of charged-particle tracks, calorimeter clusters, and muon tracks. These building bricks, or “elements”, of the particle-flow event reconstruction must therefore be delivered with a high efficiency and a low fake rate, even in high-density environments. These constraints led to the development of advanced tracking and clustering algorithms. Their concepts are briefly presented in this section.

2.1 Iterative Tracking

The momentum of charged hadrons is measured in the tracker with a resolution vastly superior to that of the calorimeters for p_T up to several hundreds of GeV/c. Furthermore, the tracker provides, before any deviation by the magnetic field during the propagation to the calorimeters, a precise measurement of the charged-particle direction at the production vertex. Because about two thirds of the energy of a jet is on average carried by charged particles, the tracker thus stands out as the cornerstone of the particle-flow event reconstruction.

Since each charged hadron missed by the tracking algorithm would be solely (if at all) detected by the calorimeters, hence with reduced efficiency, largely degraded energy resolution and biased direction, the tracking efficiency must be as close to 100% as possible. Somewhat in contradiction with this constraint, the tracking fake rate must be kept small because fake tracks, with a randomly distributed momentum, would lead to potentially large energy excesses.

An iterative-tracking strategy [2] was adopted to achieve both high efficiency and low fake rate. First, tracks are seeded and reconstructed with very tight criteria, leading to a moderate tracking efficiency, but a negligibly small fake rate. The next steps proceed by removing hits unambiguously assigned to the tracks found in the previous iteration, and by progressively loosening track seeding criteria. The softer seeding criteria increase the tracking efficiency, while the hit removal allows the fake rate to be kept low due to the reduced combinatorics. In the first three iterations, tracks originating from within a thin cylinder around the beam axis are found with an efficiency of 99.5% for isolated muons in the tracker acceptance, and larger than 90% for charged hadrons in jets.

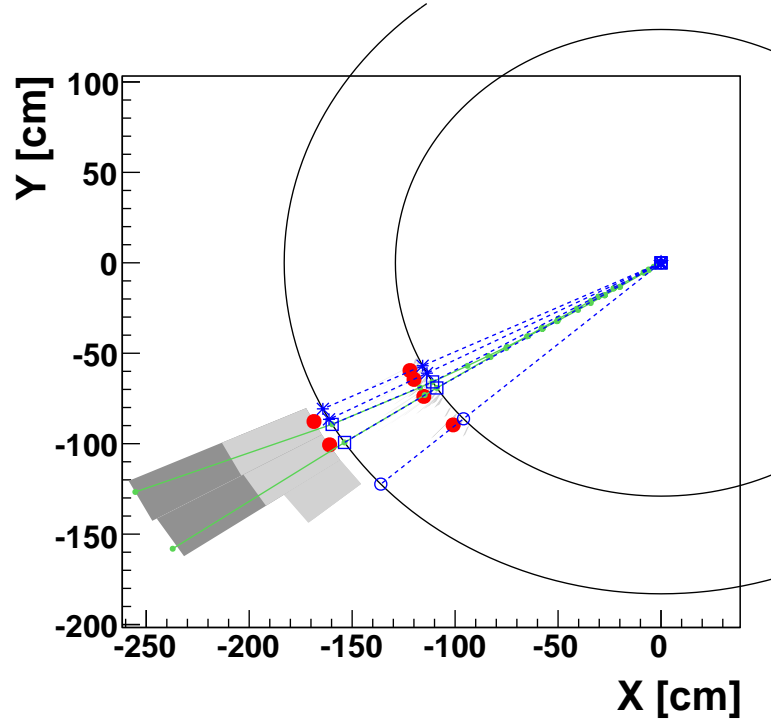
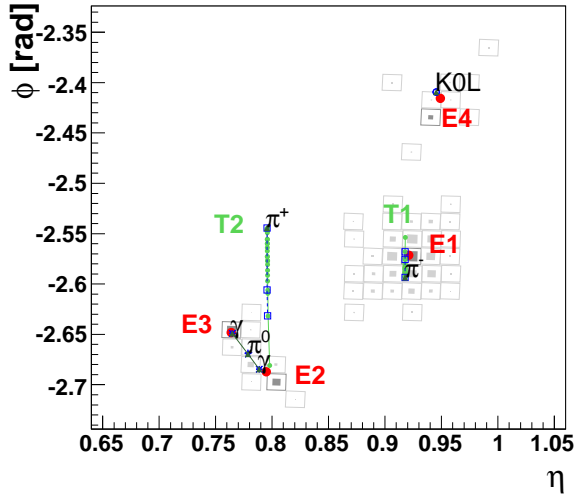
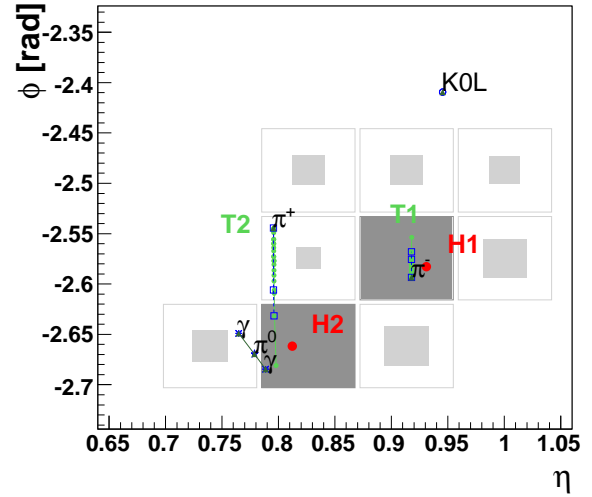
(a) The (x, y) view(b) The (η, ϕ) view on ECAL(c) The (η, ϕ) view on HCAL

Figure 1: An event display of a simple hadronic jet in the (x, y) view (a) and in the (η, ϕ) view, where η stands for pseudo-rapidity and ϕ for the azimuthal angle, on the ECAL surface (b) and the HCAL surface (c). (These two surfaces are represented as two circles centred around the interaction point in the first view.) The K_L^0 , the π^- and the two photons from the π^0 decay are detected as four well separated ECAL clusters (b). The π^+ leaves no energy in the ECAL. The two charged pions are reconstructed as charged-particle tracks, appearing as vertical solid lines in the (η, ϕ) views and circular arcs in the (x, y) view. These tracks point towards two HCAL clusters (c). In all three views, the cluster positions are represented by dots, the simulated particles by dashed lines, and the position of their impact on the calorimeter surfaces by various open markers.

The fourth and fifth iterations have relaxed constraints on the origin vertex, which allows the reconstruction of secondary charged particles originating from photon conversions and nuclear interactions in the tracker material, and from the decay of long-lived particles such as K_S^0 's or Λ 's. With this iterative technique, charged particles with as little as three hits, a p_T as small as 150 MeV/c and an origin vertex more than 50 cm away from the beam axis, are reconstructed with a fake rate of the order of a per cent.

2.2 Calorimeter Clustering

The purpose of a clustering algorithm in the calorimeters is at least fourfold: (i) detect and measure the energy and direction of stable neutral particles such as photons and neutral hadrons; (ii) separate these neutral particles from energy deposits from charged hadrons; (iii) reconstruct and identify electrons and all accompanying Bremsstrahlung photons; and (iv) help the energy measurement of charged hadrons for which the track parameters were not determined accurately, which is the case for low-quality, or high- p_T tracks.

A specific clustering algorithm has therefore been developed for the particle-flow event reconstruction, with the aim of a high detection efficiency even for low-energy particles, and towards a separation of close energy deposits, as illustrated in Fig. 1. The clustering is performed separately in each sub-detector: ECAL barrel, ECAL endcap, HCAL barrel, HCAL endcap, PS first layer and PS second layer. In the HF, no clustering is performed so far, so that each cell gives rise to one cluster.

The algorithm consists of three steps. First, “cluster seeds” are identified as local calorimeter-cell energy maxima above a given energy. Second, “topological clusters” are grown from the seeds by aggregating cells with at least one side in common with a cell already in the cluster, and with an energy in excess of a given threshold. These thresholds represent two standard deviations of the electronics noise in the ECAL (*i.e.* 80 MeV in the barrel and up to 300 MeV in the end-caps) and amount to 800 MeV in the HCAL. A topological cluster gives rise to as many “particle-flow clusters” as seeds. The calorimeter granularity is exploited by sharing the energy of each cell among all particle-flow clusters according to the cell-cluster distance, with an iterative determination of the cluster energies and positions. In the simple-jet example of Section 1, the resulting clusters, four of which are resolved from two topological clusters (one in ECAL and one in HCAL), are represented by dots in Fig. 1.

3 Link algorithm

A given particle is, in general, expected to give rise to several particle-flow elements in the various CMS sub-detectors: one charged-particle track, and/or several calorimeter clusters, and/or one muon track. The event display of Fig. 1 exemplifies most of the possible cases.

These elements must therefore be somehow connected to each other by a link algorithm to fully reconstruct each single particle, while getting rid of any possible double counting from different detectors. The link algorithm is tentatively performed for each pair of elements in the event and defines a distance between any two linked elements to quantify the quality of the link. The algorithm then produces “blocks” of elements linked directly or indirectly. Thanks to the granularity of the CMS detectors, blocks typically contain only one, two or three elements, and constitute simple inputs for the particle reconstruction and identification algorithm. The smallness of the blocks ensures the algorithm performance to be essentially independent of the event complexity. For example, jets as simple as that of Fig. 1, made of one to four blocks, turn out to have the same energy response and resolution as regular jets found in QCD events,

which typically feature a larger number of blocks, but of the same size.

More specifically, a link between a charged-particle track and a calorimeter cluster proceeds as follows. The track is first extrapolated from its last measured hit in the tracker to (i) the two layers of the PS; (ii) the ECAL, at a depth corresponding to the expected maximum of a typical longitudinal electron shower profile; (iii) the HCAL, at a depth corresponding to one interaction length, typical of a hadron shower. The track is linked to any given cluster if the extrapolated position in the corresponding calorimeter is within the cluster boundaries. This cluster envelope can be enlarged by up to the size of a cell in each direction, to account for the presence of gaps between calorimeter cells, cracks between calorimeter modules, for the uncertainty on the position of the shower maximum, and for the effect of multiple scattering for low-momentum charged particles. The link distance is defined as the distance in the (η, φ) plane between the extrapolated track position and the cluster position.

In an attempt to collect the energy of all Bremsstrahlung photons emitted by electrons, tangents to the tracks are extrapolated to the ECAL from the intersection points between the track and each of the tracker layers. A cluster is linked to the track as a potential Bremsstrahlung photon if the extrapolated tangent position is within the boundaries of the cluster, as defined above.

Similarly, a link between two calorimeter clusters, *i.e.*, either between an HCAL and an ECAL cluster, or between an ECAL and a PS cluster, is established when the cluster position in the more granular calorimeter (PS or ECAL) is within the cluster envelope in the less granular calorimeter (ECAL or HCAL). This envelope can be slightly enlarged as is done for the track-cluster link. The link distance is also defined in the (η, φ) plane as the distance between the two cluster positions.

Finally, a link between a charged-particle track in the tracker and a muon track in the muon system is established (and is called a global muon) when a global fit between the two tracks returns an acceptable χ^2 [3]. When several global muons can be fit with a given muon track and several tracker tracks, only the global muon that returns the smallest χ^2 is retained. This χ^2 defines the link distance in that case.

4 Particle reconstruction and identification

The reconstruction and identification of a set of particles from each block of elements is finally performed by the particle-flow algorithm. The resulting list of reconstructed particles constitutes a global description of each event, available for subsequent physics analysis.

4.1 Description of the algorithm

For each block, the algorithm proceeds as follows. First, each global muon gives rise to a “particle-flow muon”, if its combined momentum is compatible with that determined from the sole tracker within three standard deviations. The corresponding track is removed from the block. An estimate of the energy deposited in the HCAL (ECAL), used at a later stage in the algorithm, was measured with cosmic rays to be 3 (0.5) GeV, with an uncertainty of $\pm 100\%$.

Electron reconstruction and identification follows. Because it does not noticeably affects the performance in terms of jet, tau and E_T^{miss} reconstruction, only a brief account of this aspect of the algorithm is given here. Each track of the block is submitted to a pre-identification stage which exploits the tracker as a pre-shower: electrons tend to give rise to short tracks, and to lose energy by Bremsstrahlung in the tracker layers on their way to the calorimeter. Pre-identified electron tracks are refit with a Gaussian-Sum Filter [4] in an attempt to follow their trajectories

all the way to the ECAL. A final identification is performed with a combination of a number of tracking and calorimetric variables. Each identified electron gives rise to a “particle-flow electron”. The corresponding track and ECAL clusters (including all ECAL clusters identified as Bremsstrahlung photons) are removed from further processing of the block.

Tighter quality criteria are applied to the remaining tracks: it is required that the relative uncertainty on the measured p_T be smaller than the relative calorimetric energy resolution expected for charged hadrons, obtained by a calibration procedure described in Section 4.2. In hadronic jets, 0.2% of the tracks are rejected by this requirement. While about 90% of those are fake tracks, the energy of the 10% tracks originating from real particles is not lost for the particle-flow reconstruction, as it is measured independently with more precision, in this case, by the calorimeters. The remaining elements may give rise to charged hadrons, photons or neutral hadrons, and more rarely to additional muons.

A track can be directly connected to a number of ECAL and HCAL clusters. The detection of the neutral particles in the block (photons and neutral hadrons) involves a comparison between the momentum of the tracks and the energy detected in the calorimeters. For this comparison to be reliable, the ECAL and HCAL cluster energies, from which the expected muon energy deposits are subtracted, must undergo the calibration procedure of Section 4.2. Several tracks can be linked to the same HCAL cluster, in which case the sum of their momenta is compared to the calibrated calorimetric energy. On the other hand, if a track is linked to several HCAL clusters, only the link to the closest cluster is kept for the comparison. A track can also be linked to more than one ECAL clusters, and the link to the closest cluster is kept too. The possible additional ECAL clusters might come from hadronic shower fluctuations, in which case the links ought to be preserved to avoid double counting of the hadron energy. Conversely, if these ECAL clusters arise from overlapping photons, the links ought to be ignored to allow the photon detection. To take this decision, the ECAL clusters connected to any of the tracks under consideration are first ordered according to their distance to the closest track. The ordered list is then scanned and the corresponding link is kept as long as the total calibrated calorimetric energy (from the HCAL cluster, if any, and from all ECAL clusters considered at this point) remains smaller than the total charged-particle momentum.

In rare cases, the total calibrated calorimetric energy is still smaller than the total track momentum by a large amount. When the difference is larger than three standard deviations, a relaxed search for muons and for fake tracks is performed. First, all global muons, not already selected by the algorithm and for which an estimate of the momentum exists with a precision better than 25%, are treated as particle-flow muons. The redundancy of the measurements in the tracker and the calorimeters thus allows a few more muons to be found without increasing the fake-muon rate. This redundancy is further exploited by progressively removing tracks from the block, ordered according to their measured p_T uncertainty. The process stops either when all tracks with an p_T uncertainty in excess of 1 GeV/c have been examined, or when the removal of a track would render the total track momentum smaller than the calibrated calorimetric energy. Less than 0.3 per mil of the tracks are concerned by this procedure.

Each of the remaining tracks in the block gives rise to a “particle-flow charged hadron”, the momentum and energy of which are taken directly from the track momentum, under the charged pion mass hypothesis. If the calibrated calorimetric energy is compatible with the track momentum within measurements uncertainties, the charged-hadron momenta are redefined by a fit of the measurements in the tracker and the calorimeters, which reduces to a weighted average if only one track is present. This combination is relevant at very high energies and/or large pseudo-rapidities, for which the track parameters are measured with degraded resolutions.

On the other hand, it may well be that the calibrated energy of the closest ECAL and HCAL clusters linked to the track(s) be significantly larger than the total associated charged-particle momentum. If the relative energy excess is found to be larger than the expected calorimeter energy resolution (quadratically augmented by the uncertainty due to the presence of muons in the block), it gives rise to a “particle-flow photon”, and possibly to a “particle-flow neutral hadron”. Specifically, if the excess is larger than the total ECAL energy, a photon is created with this ECAL energy and a neutral-hadron is created with the remaining part of the excess, calibrated as described in Section 4.2. Otherwise, the uncalibrated excess gives rise only to a photon. The precedence given in the ECAL to photons over neutral hadrons is justified by the observation that, in jets, 25% of the jet energy is carried by photons, while neutral hadrons leave only 3% of the jet energy in the ECAL. This fraction is reduced by one order of magnitude for taus, for which decays to final states with neutral hadrons are Cabbibo-suppressed to a branching ratio of about a per cent.

The remaining ECAL and HCAL clusters, either originally not linked to any track or for which the link was disabled, give rise to particle-flow photons and particle-flow neutral hadrons, respectively. The neutral-hadron energies are determined by applying the calibration procedure to the HCAL clusters only.

4.2 Cluster calibration

The ECAL is already meant to be calibrated for photons and electrons [1]. The residual corrections, *e.g.*, for thresholds or specificities of the clustering algorithm, amount to a couple per cent only. These corrections can be safely obtained with simulated photons and are therefore not described here.

Hadrons deposit energy, in general, in both ECAL and HCAL. The former is calibrated for photons, and the latter to 50 GeV pions not interacting in ECAL. Because the HCAL response to hadrons is nonlinear and because the ECAL response to hadrons is different from the response to photons, the ECAL and HCAL cluster energies need to be substantially recalibrated to get an estimate of the true hadronic energy deposits in ECAL and HCAL. It is worth repeating, however, that this calibration affects mostly neutral hadrons, which represent only 10% of the measured event energy. The latter is therefore expected to be modified, on average, by only a few per cent by the calibration procedure.

In the particle-flow algorithm, the calorimetric energy linked to charged hadrons is calibrated by the following calibration function

$$E_{\text{calib}} = a + b(E, \eta)E_{\text{ECAL}} + c(E, \eta)E_{\text{HCAL}} \quad (1)$$

where E_{ECAL} and E_{HCAL} are the energies measured in ECAL and HCAL, η the pseudo-rapidity of the HCAL cluster, and E an estimate of the true energy, chosen to be either the total charged-particle momentum or the total (un-calibrated) calorimetric energy, whichever is larger.

For the purpose of this note, a sample of simulated single hadrons was used to determine the calibration coefficients a , b and c . For a given of value a , the coefficients b and c are obtained by minimizing, in each bin of E , the following χ^2 :

$$\chi^2 = \sum_{i=1}^N \frac{(E_{\text{calib}}^i - E^i)^2}{\sigma_i^2(E_{\text{calib}}^i)}, \quad (2)$$

where E^i and σ_i are the true energy and the expected calorimetric energy resolution of the i^{th} single hadron, and where the sum extends over all events, separately (a) in the barrel and

the end-cap regions of the calorimeter; and (b) for hadrons leaving energy either solely in the HCAL, or solely in the ECAL, or in both calorimeters.

The energy resolution $\sigma_i(E)$ is determined iteratively as the Gaussian sigma of the $(E_{\text{calib}} - E)$ distributions in each bin of true energy, and is shown in Fig. 2a as a function of E . The coefficient a is obtained iteratively to minimize the dependence of b and c on E , so as to cancel systematic uncertainties for neutral hadrons, for which no accurate estimate of the true energy in real data is available. This coefficient is found to amount to about 3 GeV, and can be understood as a constant correction for the thresholds applied to the calorimetric cell energies in the clustering algorithm. With this choice, the dependence on E of b and c , for hadrons leaving energy in HCAL (barrel region) is shown in Fig. 2b. The larger c coefficient for the hadrons that leave energy in ECAL is meant at compensating the energy lost in the dead material between ECAL and HCAL.

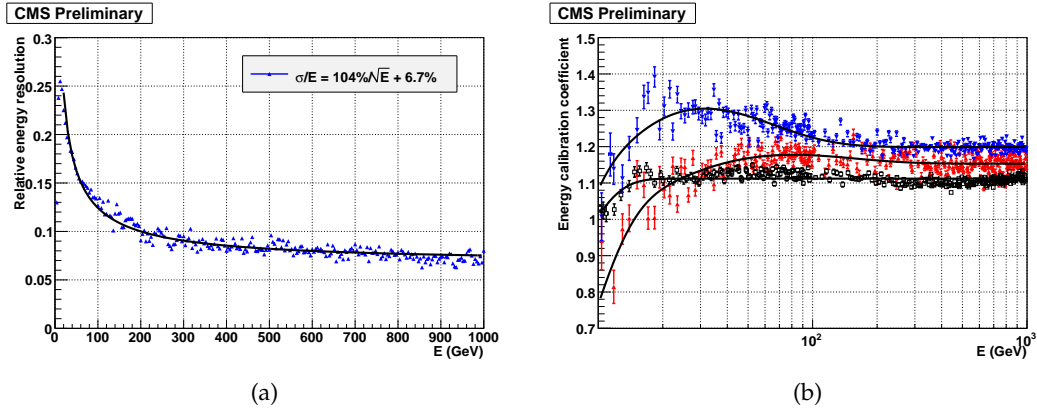


Figure 2: Energy resolution σ/E as a function of the true hadron energy E (a). Calibration coefficients as a function of E (b), for hadrons depositing energy in HCAL barrel only (open squares), and for hadrons depositing energy in both ECAL and HCAL barrel, for ECAL (downward triangles) and HCAL (upward triangles). The smooth curves are obtained with a fit of the data points to *ad-hoc* functions of E , used in the particle-flow algorithm.

The calibration coefficients a , b and c , and the parameterized resolution σ will be obtained prior to the start of the collision data taking from combined ECAL-HCAL test-beam data. A preliminary analysis of these data confirm that the values deduced from the simulation are upheld by real data. A cross-check with collision data will be done with a sample of isolated charged-particle tracks, not identified as electrons or muons, where the true energy E will be estimated from the track momentum directly. The details of the isolated-track selection deemed adequate for this cross-check are currently under study.

5 Performance of the Particle-Flow Event Reconstruction

The following samples of events were generated with PYTHIA [5] and processed through the GEANT-4-based simulation [6] (hereafter called full simulation) of the detector:

- QCD multijets events, with a flattened \hat{p}_T spectrum from 15 GeV/c to 1.5 TeV/c, for jet and E_T^{miss} performance studies;
- Semi-leptonic $t\bar{t}$ events, for E_T^{miss} performance studies;
- $Z \rightarrow \tau\tau$ with hadronic τ decays, for τ reconstruction performance studies.

A fast but detailed simulation of the detector [1] was used for some of the systematic studies. The performance of the particle-flow event reconstruction for jets, taus and E_T^{miss} have been evaluated so far for momentum transfer below 1 TeV/c, and in the tracker acceptance for jets and taus, where particle-flow techniques are expected to bring most. Dedicated developments in the rest of the parameter space are underway.

5.1 Performance with Jets

As mentioned previously, the typical jet energy fractions carried by charged particles, photons and neutral hadrons are 65%, 25% and 10% respectively. These fractions ensure that 90% of the jet energy can be reconstructed with good precision by the particle-flow algorithm, both in value and direction, while only 10% of the energy is affected by the poor hadron calorimeter resolution and by calibration corrections of the order of 10 to 20%. As a natural consequence, it is expected that jets made of reconstructed particles be much closer to jets made of Monte-Carlo-generated particles than jets made from the sole calorimeter information, in energy, direction and content. It is the purpose of this section to quantify this statement.

5.1.1 Jet reconstruction

Jets are reconstructed from the QCD-multijet event sample with the iterative-cone algorithm [3] with a cone size of 0.5 in the (η, φ) plane, from several types of inputs.

- All generated stable particles, except for neutrinos, give rise to so-called “gen-jets”;
- All particles reconstructed with the particle-flow algorithm, without distinction of type and without any energy threshold, are clustered into “particle-flow jets”;
- Calorimeter towers, made of the energy contained in each HCAL cells and the 25 underlying ECAL crystals, give birth, if this energy exceeds 1 GeV, to “calo-jets”.

The reconstructed jets are then matched to the closest gen-jet in the (η, φ) plane. For particle-flow jets, it is requested that the distance in the (η, φ) plane between the reconstructed jet and the matched generated jet be smaller than 0.1. The matching jet efficiency, *i.e.*, the fraction of generated jets that give rise to a matched reconstructed jet, and the mismatched jet rate, *i.e.*, the fraction of reconstructed jets that do not have a matched generated jet, are shown in Fig. 3 as a function of the jet p_T , in the barrel and the end-caps. For the jet p_T 's studied in this note (in excess of 20 GeV/c), a efficiency larger than 80% is obtained. The 100% plateau is reached above 40 GeV/c, at which point the mismatched jet rate is negligible.

The matching efficiency and mismatched jet rate depend substantially on the matching criterion. For example, Fig. 4 displays the same quantities for a matching distance of 0.2 instead of 0.1, here for the barrel only. For this value, the efficiencies and mismatching rate obtained for calo-jets become similar to those observed for particle-flow jets with a distance of 0.1. A matching distance of 0.2 is thus chosen for calo-jets for a fair comparison to particle-flow jets, together with a minimum generated jet p_T of 15 GeV/c to select calo-jets with an efficiency larger than 50%. Particle-flow jets can, however, be used in analysis down to p_T 's as small as 5 GeV/c. Finally, the matching distance is increased to 0.5 in Fig. 5. The combined use of tracks and calorimeter clusters allows the efficiency for particle-flow jets to reach 100% independently of p_T . The sole use of the calorimeter information is still affected at low p_T by the calorimeter tower energy thresholds.

The distributions of the ratio $(p_T^{\text{rec}} - p_T^{\text{gen}})/p_T^{\text{gen}}$, where “rec” and “gen” hold for reconstructed and generated jets, respectively, are exemplified in Fig. 6 for two different p_T^{gen} bins.

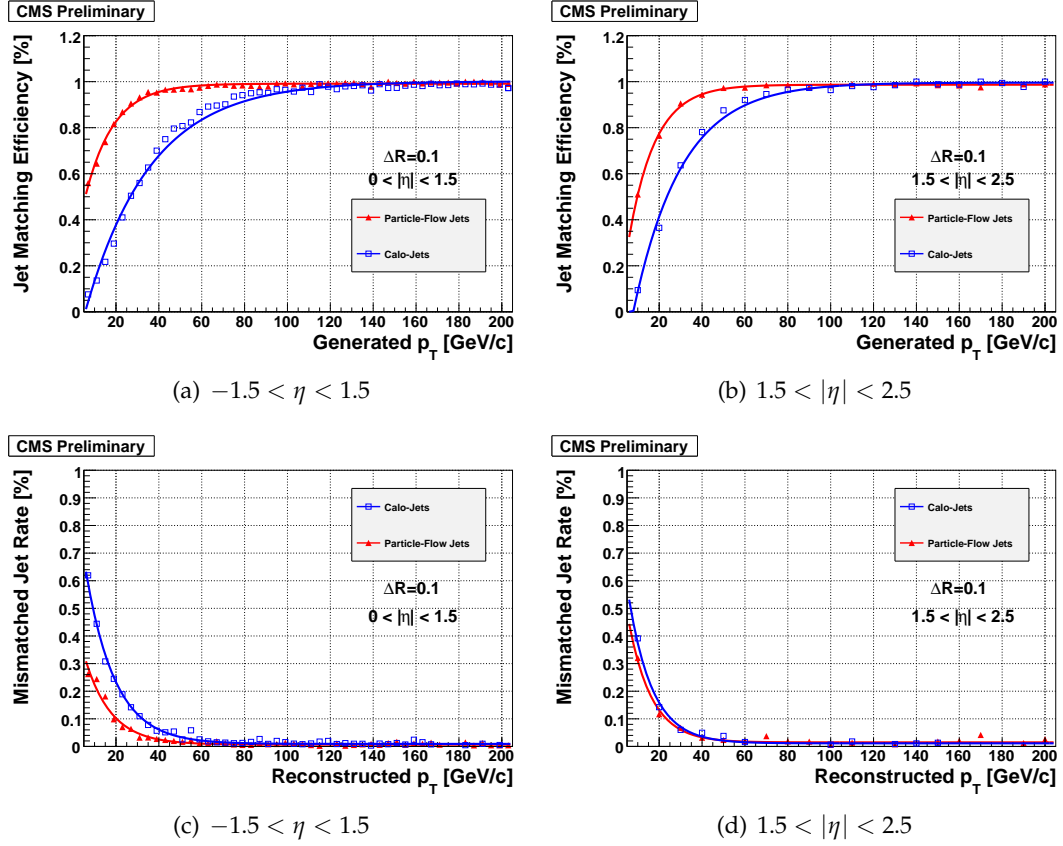


Figure 3: Jet matching efficiency (a,b) and mismatched jet rate (c,d) as a function of the jet p_T , as obtained for calo-jets (open squares) and particle-flow jets (triangles) pointing to the barrel, $-1.5 < \eta < 1.5$ (a,c) and to the end-caps, $1.5 < |\eta| < 2.5$ (b,d), with a matching distance of 0.1. Efficiencies and fake rates are fit to exponential functions of p_T .

A Gaussian is fit to these distributions. The response is defined as the mean value μ of this Gaussian, and is shown for particle-flow and calo jets in Fig. 7. The particle-flow-jet response benefits from the reconstruction of all particles in the event from a combination of all CMS sub-detectors, which ensures that little energy is lost over the whole acceptance. The systematic uncertainty related to the residual response correction (Section 5.1.2) is reduced accordingly.

Finally, the jet-direction resolutions are displayed in Fig. 8 both in η and ϕ , for the barrel, as a function of p_T . The jet-direction precision is relevant when it comes to measuring jet-jet invariant masses, in particular for top-quark or exotic-particle decay reconstruction. The achieved resolutions benefit from the clustering algorithm, which exploits the granularity of the ECAL. Furthermore, the charged particles are bent by the magnetic field, leading to a degradation of the ϕ resolution of the calo-jets. In the case of particle flow, the charged particles are reconstructed at their production vertex, hence similar resolutions along the η and ϕ directions.

5.1.2 Jet Energy Resolutions

The jet-energy resolution is, like the jet-direction resolution, an important ingredient for the determination of the di-jet invariant masses. As shown in Fig. 7, particle-flow jets have an energy scale already very close to unity, hence would need only small (if any) residual corrections. Several procedures exist to bring the energy scale of calo-jets up to unity.

The default jet-energy correction [7], based on large samples of simulated QCD events, brings

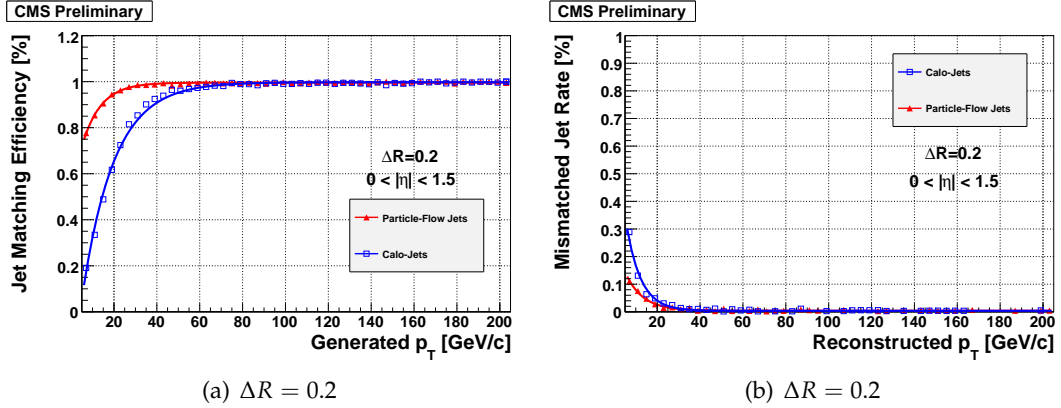


Figure 4: Jet matching efficiency (a) and mismatched jet rate (b), as obtained for calo-jets (open squares) and particle-flow jets (triangles) in the barrel, with a matching distance of 0.2. Efficiencies and fake rates are fit to exponential functions of p_T .

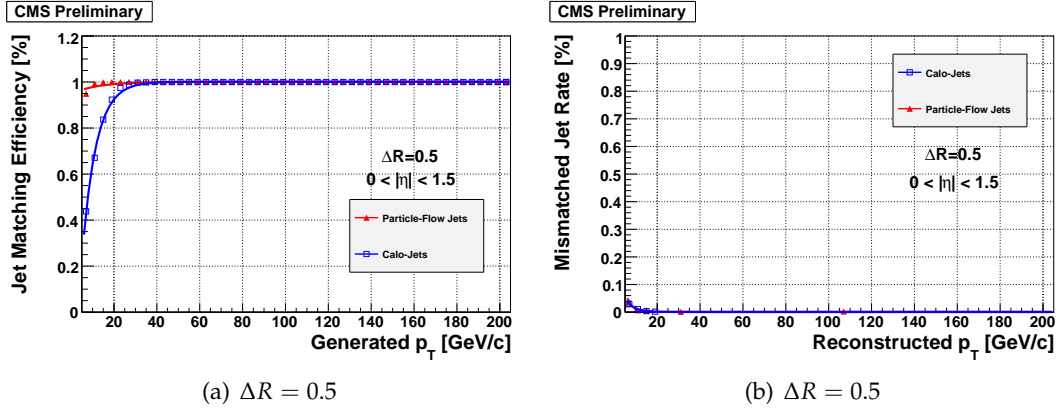


Figure 5: Jet matching efficiency (a) and mismatched jet rate (b), as obtained for calo-jets (open squares) and particle-flow jets (triangles) in the barrel, with a matching distance of 0.5. Efficiencies and fake rates are fit to exponential functions of p_T .

the jet energy back to that of the matched generated jet, on average, for a given p_T and pseudo-rapidity. It consists of two sequential steps: the relative correction removes the η dependence of the energy response, and the absolute correction restores the response to unity. More specifically, distributions similar to those presented in Fig. 6 are fit by a Gaussian in each (η, p_T) bin, with central value μ and width σ . All jet energies in this bin are corrected by the factor $1/(1 + \mu)$, for the energy response to reach unity. The relative energy resolution of the corrected jets (still called calo-jets in the following) is simply $\sigma/(1 + \mu)$. Due to the non-linear response of the calorimeter jets, the correction improves significantly the relative energy resolution, by correcting much more the low-response jets than the high-response jets.

At the time of writing, such corrections were not available for particle-flow jets with the latest version of the particle-flow reconstruction software. The particle-flow jet-energy resolutions, obtained by dividing the Gaussian σ by the average jet response in each p_T bin, are therefore shown without any corrections in Fig. 9, and compared to the fully corrected calo-jets. These corrections, whenever applied, will only improve the jet-energy resolutions by a residual cooling effect and by the removal of a residual η dependence of the jet response. The jet-energy resolutions presented here for particle-flow jets are therefore conservative in this respect.

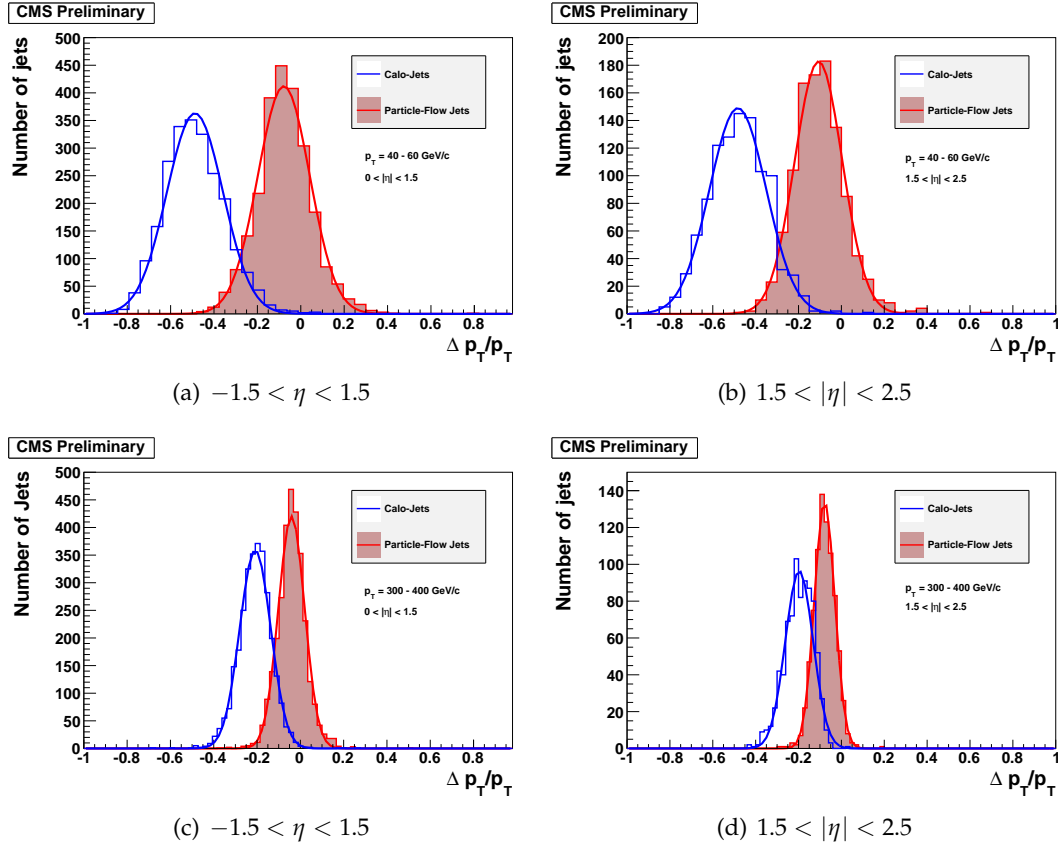


Figure 6: Distributions of $(p_T^{\text{rec}} - p_T^{\text{gen}})/p_T^{\text{gen}}$ for p_T^{gen} between 40 and 60 GeV/c (a,b) and between 300 and 400 GeV/c (c,d), as obtained from calo-jets (open histogram) and from particle-flow jets (solid histogram) pointing to the barrel (a,c) and to the end-caps (b,d). A Gaussian is fit to all distributions, to determine the response and the resolution.

5.1.3 Systematic uncertainties

Six kinds of systematic effects were studied so far: (i) the effect of a poor modelling of the calorimeter response to hadrons; (ii) the effect of a poor parameterization of the calorimeter energy resolution; (iii) the effect of a poor knowledge of the noise in the calorimeter cells, hence of the change of energy thresholds; (iv) the effect of a global reduction of the tracking efficiency; (v) the effect of a poor modelling of the tracking efficiency in high-energy, dense, jets; and (vi) the effect of the flavour of the jet-initiating parton. In all these cases, the relevant parameter values were purposely scanned over ranges much larger than the resolution expected on these parameters at the beginning or after a few months of collision-data taking. The result of these studies is presented here on the jet response in the barrel, but effects of the same order are seen in the end-caps. No sizeable effects on the jet resolution were observed in any of these cases.

To estimate the consequence of the modelling of the calorimeter response to hadrons, the energy correction needed for hadrons in the data was assumed to be different from that derived from the simulation by $\pm 50\%$, both in the ECAL and the HCAL. Looking at Fig. 2, it would correspond to correct calorimeter clusters with an energy of 100 GeV by 10% or 30% instead of 20%. This range is extreme in the sense that the available combined ECAL–HCAL test-beam data will reduce the uncertainty on the correction to a few percent well prior to the first collision data. The consequence of such an over-conservative change on the particle-flow jet energy scale is shown in Fig. 10 to be limited: the scale changes only by $\pm 3\%$ at large p_T , and $\pm 1\%$ for

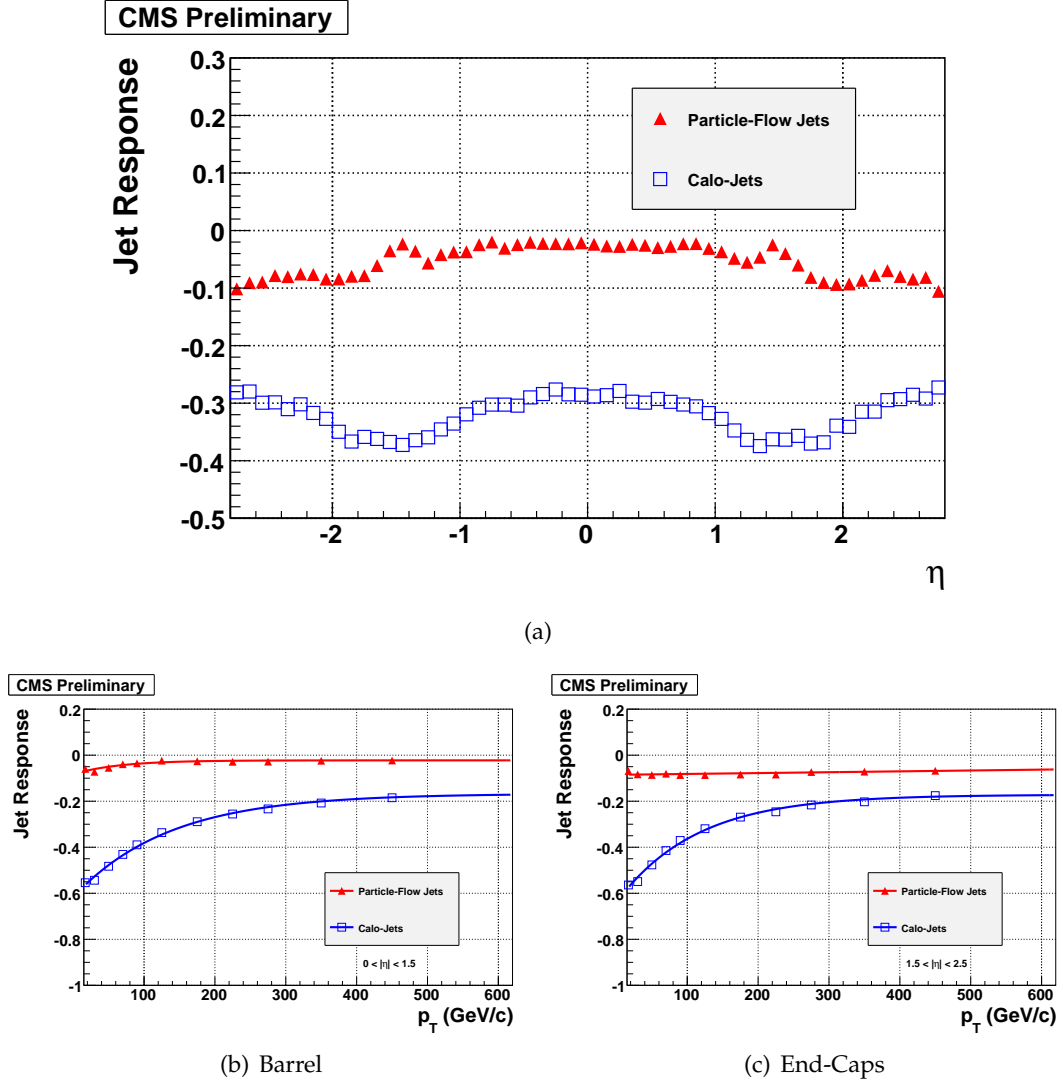


Figure 7: Jet Response as a function of η integrated over all p_T 's below 750 GeV/c (a) and as a function of p_T , in the barrel (b) and in the end-caps (c). The response curves are fit with exponential functions of p_T .

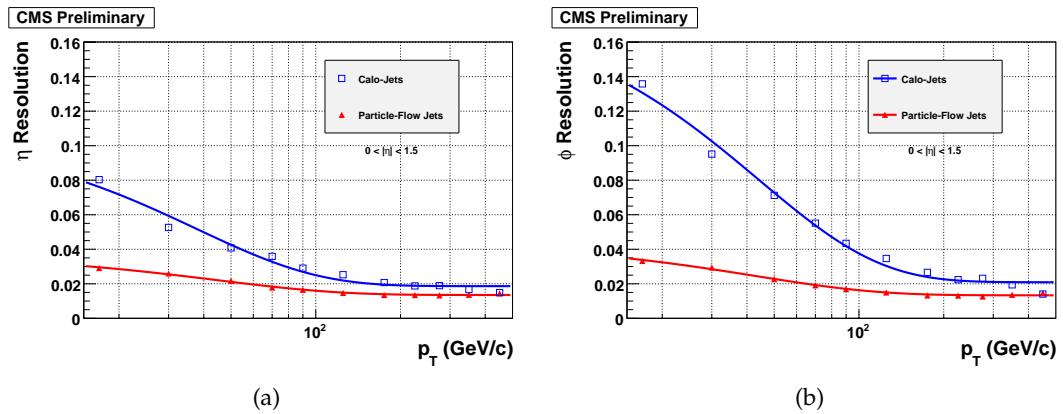


Figure 8: Jet η (a) and ϕ (b) resolutions (RMS) as a function of p_T in the barrel. The resolution curves are fit with exponential functions of p_T .

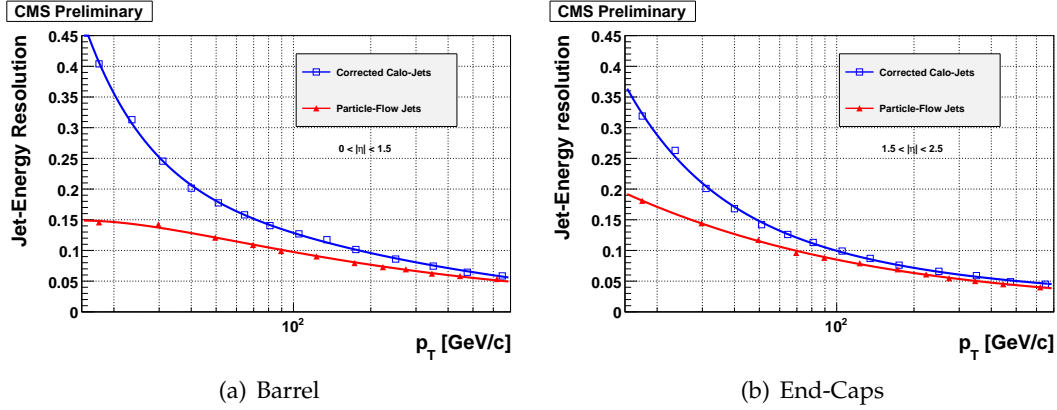


Figure 9: Jet-energy resolutions as a function of p_T for corrected calo-jets (open squares) and for particle-flow jets (upwards triangles) in the barrel (a) and in the end-caps (b). The resolution curves are fit to the sum of a constant term, a stochastic term and a noise term.

p_T 's below 100 GeV/c. This observation results from the fact that only 10% of the jet energy is carried by neutral hadrons, hence subject to energy corrections.

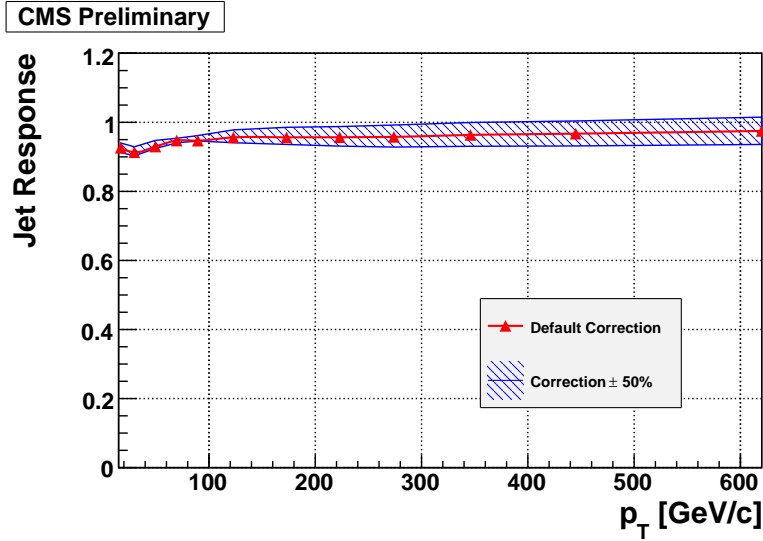


Figure 10: Response variation for particle-flow jets when the energy corrections needed for hadrons are wrong by $\pm 50\%$ (presented here in the barrel).

Similarly, and although the combined ECAL–HCAL test-beam data will allow an uncertainty smaller by an order of magnitude to be obtained, the parameterization of the calorimeter energy resolution shown in Fig. 2a was drastically changed from $50\%/\sqrt{E} \oplus 3\%$ to $150\%/\sqrt{E} \oplus 9\%$. This parameterization is used for the identification of fake tracks, muons, merged photons and merged neutral hadrons. It is therefore important to understand its impact on the particle-flow performance. The consequence of this change on the jet response is shown in Figs. 11a and b to be smaller than half a per cent over the full p_T spectrum.

To account for a uniform modification by a factor three of the electronics noise in the calorimeters, the cell thresholds were changed from 40 to 120 MeV in the ECAL and from 400 MeV to 1.2 GeV in the HCAL. These threshold variations ought to be accompanied by the corresponding calibration coefficient adjustment, but their impact, shown in Figs. 11c and d, was conservatively estimated with the same calibration coefficients. Changes in jet response of the

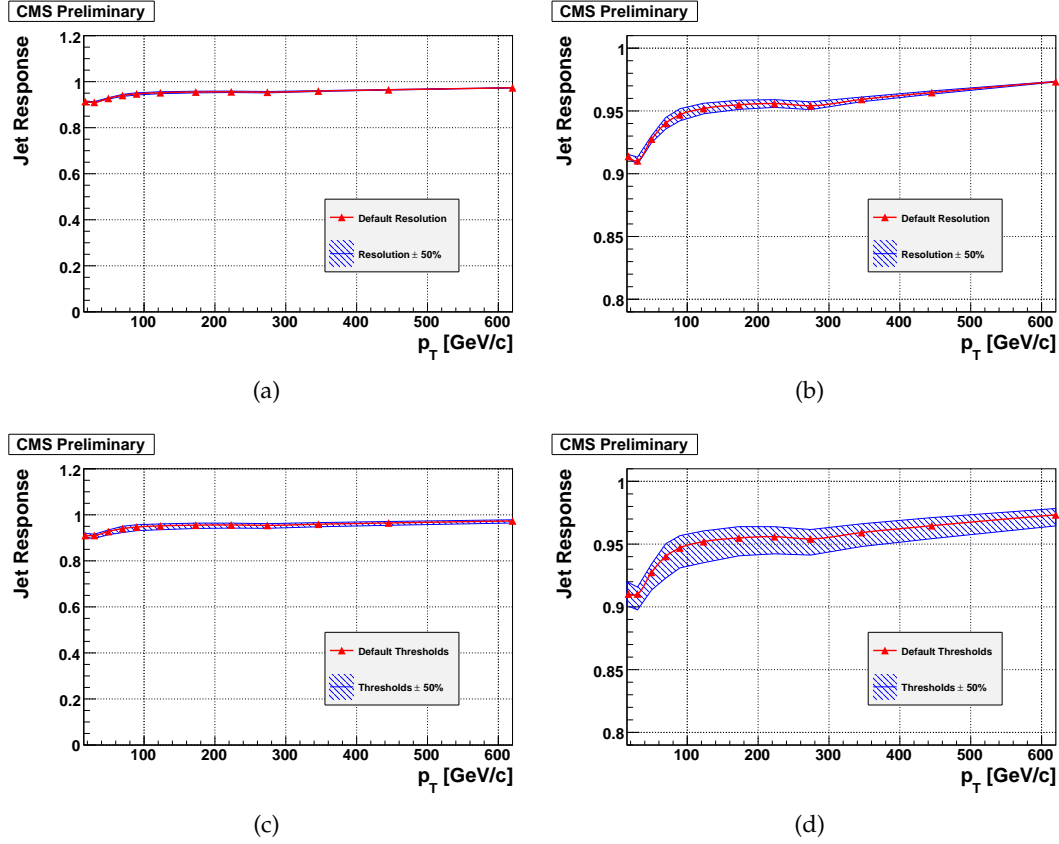


Figure 11: Response variation for particle-flow jets when the parameterized energy resolution is changed by $\pm 50\%$ (a), and when the thresholds are changed from 40 to 120 MeV in the ECAL and from 400 MeV to 1.2 GeV in the HCAL (c), presented here in the barrel. In (b) and (d), a zoom of the same figures is shown for a better appreciation of the differences.

order of $\pm 1\%$ are expected in that case, especially visible for the smallest p_T 's.

The previous three figures demonstrate the intrinsic robustness of the particle-flow event reconstruction with respect to large uncertainties in the calorimeter behaviour. Similarly, the effect of a global reduction of the tracking efficiency by 5% is displayed in Fig. 12a and b. The tracking efficiency might be smaller than expected for a variety of reasons (tracker misalignment, noise, ...), but it will be known to a precision of a per cent very soon after the start of the collision-data taking. The reduction of less than 1% of the jet response is therefore very conservative an estimate of the corresponding systematic uncertainty. The moderate consequence of such a reduction in tracking efficiency is due to the built-in redundancy of the particle-flow event reconstruction.

The tracking efficiency in the core of high-energy, dense, jets has been observed to be substantially reduced in the full simulation [8], by up to 30% (50%) for jet p_T 's of 600 GeV/c (1 TeV/c). To determine the magnitude of the systematic bias related to the modelling of this effect, a comparison of the jet response with that obtained with the fast simulation was performed. Indeed, in the fast simulation, all charged particles that satisfy the seeding conditions are reconstructed, irrespective of the density of the surrounding environment. It can be seen from Figs. 12c and d, however, that the small jet-response difference between the fast and the full simulation does not vary with p_T , and is therefore not affected by a large reduction of tracking efficiency at high p_T . This observation results from the fact that, in the core of the jets and at very high energies,

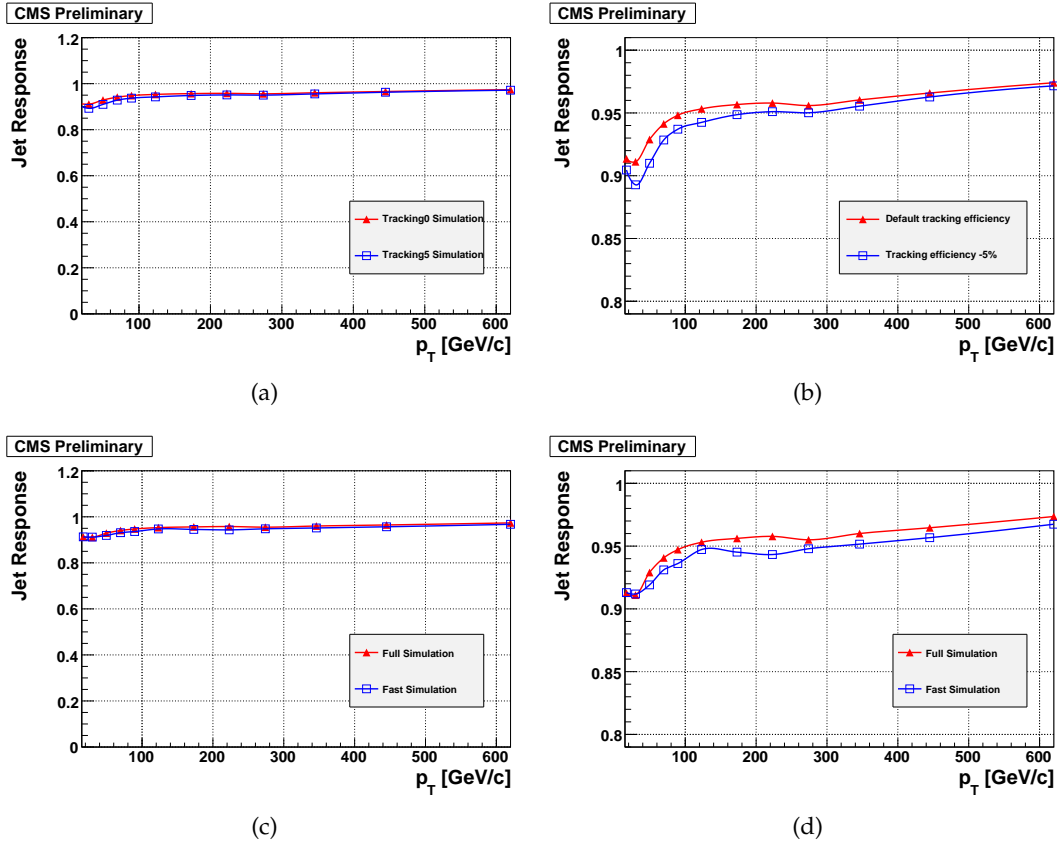


Figure 12: Response variation for particle-flow jets (a) for the default tracking efficiency (triangles) and a tracking efficiency reduced by 5% (open squares), and (b) as obtained for the full (triangles) and the fast (squares) simulations of the detector, presented here in the barrel. In (b) and (d), a zoom of the same figures is shown for a better appreciation of the differences.

the particle-flow performance are driven by those of the calorimeters.

The description of the amount and type of material in the tracker might also create systematic biases in the jet energy response. A quick look at Fig. 7a shows, however, that the energy response does not vary with the jet pseudo-rapidity over the whole barrel, while the tracker thickness varies from about 0.4 to 1.6 radiation lengths over the same range. Another check of the weak influence of the tracker material was obtained by switching on and off all interactions (photon conversions, electron Bremsstrahlung, dE/dx , multiple scattering and nuclear interactions) in the fast simulation of the tracker. No variation of the jet response larger than 1% was observed.

Finally, the effect of parton flavours was estimated by simulating jets originating from light quarks (u,d, and s quarks), c quarks, b quarks and gluons, with the fast simulation. The resulting particle-flow jet energy scales are shown in Fig. 13 as a function p_T . No differences larger than $\pm 2\%$ are observed of the whole p_T spectrum, reducing to $\pm 1\%$ above 100 GeV/c. This observation results from the fact that charged hadrons, photons and neutral hadrons are reconstructed and identified individually, with their own “calibration”, so that the jet reconstructed energies depend only weakly on their specific particle content.

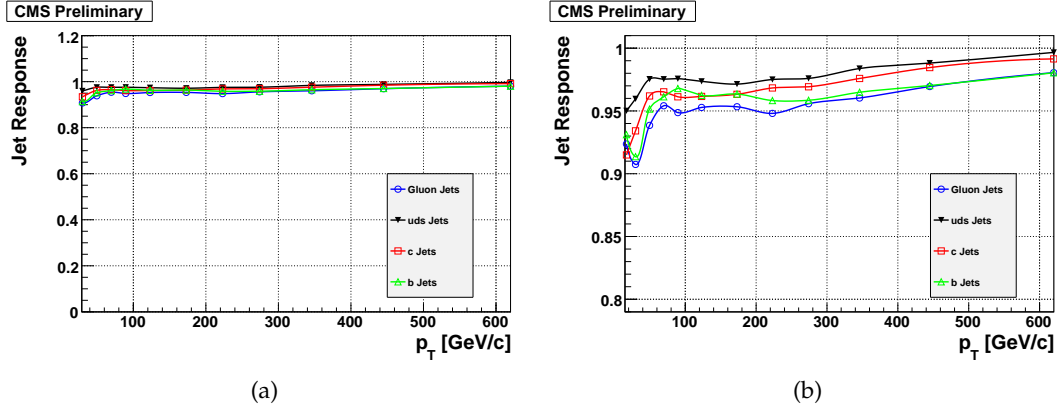


Figure 13: Response variation for particle-flow jets with different parton flavours (a): gluon jets (open circles), uds jets (solid triangles), c jets (open squares) and b jets (open triangles). In (b), a zoom of the same figure is shown for a better appreciation of the differences.

5.2 Performance with Taus

As a consequence of their large mass, taus play an important role in searches for Standard-Model and supersymmetric Higgs bosons. About two-thirds of the taus decay hadronically, most often into one or three charged hadrons (for approximately 75% and 25% of the cases respectively), possibly with a number of π^0 's immediately decaying into two photons. As the branching ratio into final states containing neutral hadrons is as small as 1%, particle-flow event reconstruction is expected to provide an excellent accuracy for the measurement of the energies and directions of all the visible tau decay products. As a matter of fact, particle-flow techniques have been used successfully in the past for tau reconstruction, even at hadron-collider experiments [9, 10].

The fully-simulated sample of $Z \rightarrow \tau\tau$ events, with both τ decaying hadronically, was used to determine the performance of the tau reconstruction. The tau identification efficiency and fake rate has been studied elsewhere in Ref. [11]. No performance change is observed with respect to this work. Tau candidates can be reconstructed using two different methods. In a traditional approach [12], each calo-jet reconstructed by the iterative cone algorithm with a radius $\Delta R = 0.5$ (Section 5.1.1) is considered a tau candidate. In the particle-flow reconstruction, the tau four-momentum is the sum of the four-momenta of all particles in a small cone of radius $\Delta R = 0.15$ around the direction of the leading particle in the jet, with p_T in excess of 0.5 GeV/c. In this way, particles from the underlying event have a minimal impact on the reconstructed tau momentum.

The stable, visible, tau decay products (thus excluding neutrinos), are clustered into tau gen-jets, which are used as a reference. For each tau gen-jet, only the closest reconstructed tau in the (η, φ) plane is considered. The reconstructed and generated transverse energies are compared in Fig. 14, both for the calorimetric and the particle-flow approaches, in the barrel and the end-caps. The particle-flow reconstruction provides an accurate estimate of the energy scale and a good energy resolution. The resolution is, however, degraded in the end-caps by the larger amount of tracker material, in which photons convert and charged hadrons undergo nuclear interactions, the products of which escape the small cone used for tau reconstruction. The determination of the tau direction also benefits from the particle-flow reconstruction, as shown in Fig. 15: unlike in the traditional reconstruction, and in addition to an improved resolution in both η and φ , the φ measurement does not exhibit any bias due to magnetic field effects.

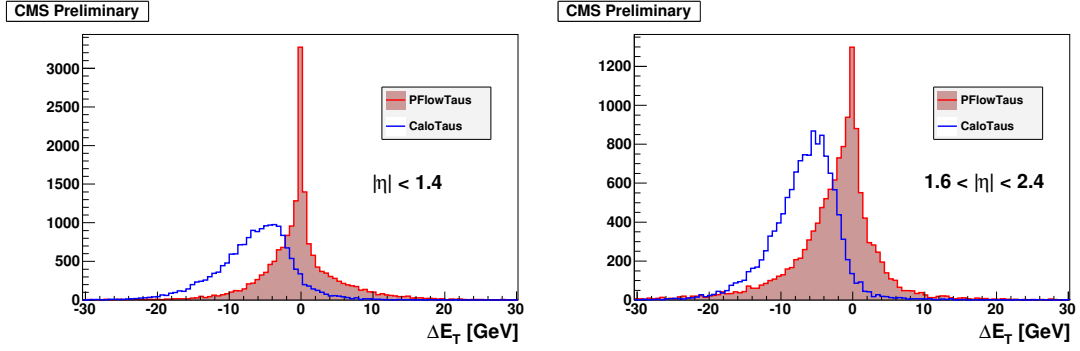


Figure 14: Resolution in E_T for hadronic τ jets from Z decay, reconstructed using the particle flow (filled), and calorimeters (hollow). Left: barrel; right: end-caps.

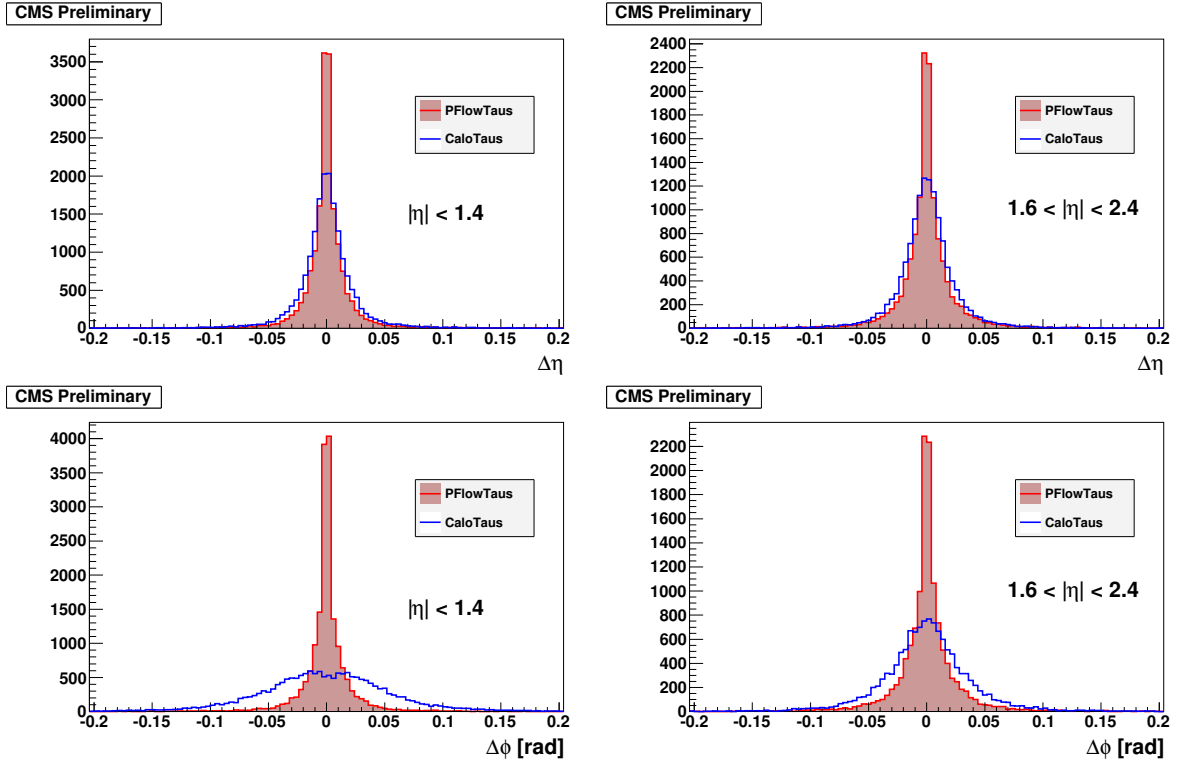


Figure 15: Angular resolution in η (top) and ϕ (bottom) for hadronic τ jets from Z decay, reconstructed using the particle flow (filled), and calorimeters (hollow). Left: barrel; right: end-caps.

5.3 Performance with Missing Transverse Energy

An accurate determination of E_T^{miss} , both for events without missing energy, typical of standard-model backgrounds, and for events with missing energy, often characterizing new physics, is a major asset for the separation of the two types of events. It is in principle simple to determine E_T^{miss} after the particle-flow event reconstruction: it merely consists in forming the transverse-momentum-vector sum over all reconstructed particles in the event and then taking the opposite of this azimuthal, momentum two-vector. The missing transverse energy is the modulus of this vector. The true E_T^{miss} is derived in a similar manner with all visible generated particles or, equivalently, with all invisible generated particles, like neutrinos and neutralinos.

The performance of E_T^{miss} , obtained from particle-flow reconstruction is compared here with that obtained from calorimeter towers. The latter is corrected for both the jet energy scale

of calorimeter jets and the presence of identified muons [13]. No a posteriori corrections are applied to E_T^{miss} from particle-flow reconstruction.

First, the total visible transverse energy, denoted $\sum E_T$, is obtained as the sum of the transverse energies of all reconstructed particles (respectively, calorimeter towers), and measures the amount of global event energy that the algorithm is able to reconstruct. The average of the relative difference between the reconstructed and the generated $\sum E_T$ is displayed in Fig. 16a as a function of the true $\sum E_T$, in the QCD multijet event sample. In these events, little missing energy is in general expected to be measured, but this result can only be achieved with a finite precision. The resolution of the measured transverse energy vector along any axis perpendicular to the beam is shown in Fig. 16b as a function of the true $\sum E_T$. The E_T^{miss} resolution converges to the calorimeter-based measurement above 3 TeV. At these energies, the performance of the particle-flow event reconstruction is indeed driven by those of the calorimeters.

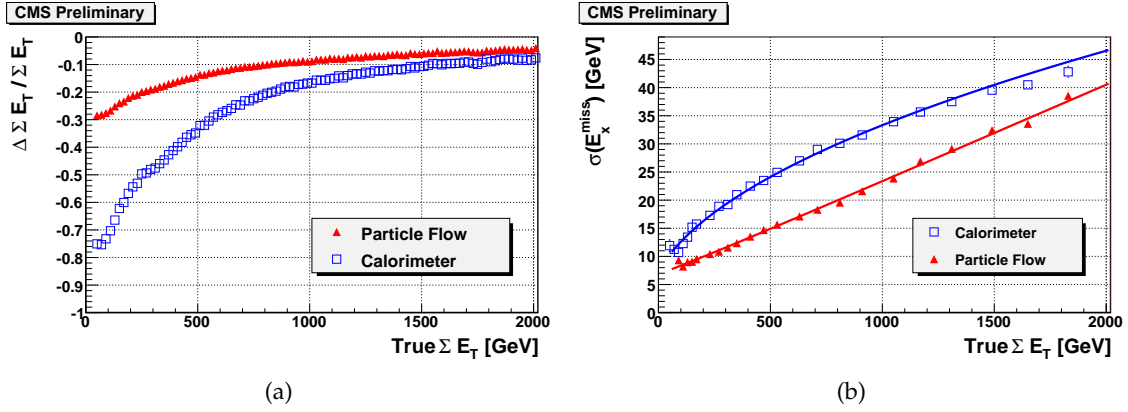


Figure 16: (a) Response, for QCD multijet events, of the total visible transverse energy of the event, defined to be $(\sum E_T^{\text{reco}} - \sum E_T^{\text{true}}) / \sum E_T^{\text{true}}$, as a function of the true total visible transverse energy of the event; (b) Resolution of the x -projection of E_T^{miss} , obtained from a Gaussian fit, versus the total true visible transverse energy of the event. The solid triangles represent quantities based on particle-flow reconstruction; the open squares represent quantities based on calorimeter reconstruction.

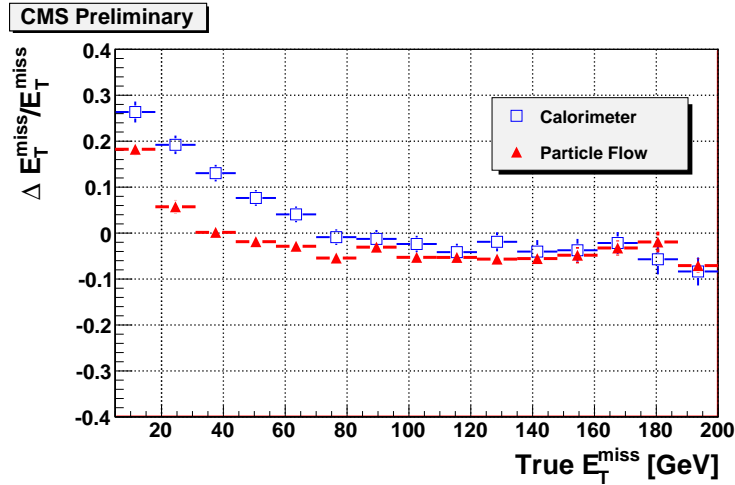


Figure 17: Distribution of $(E_{T,\text{reco}}^{\text{miss}} - E_{T,\text{true}}^{\text{miss}}) / E_{T,\text{true}}^{\text{miss}}$ as a function of the $E_{T,\text{true}}^{\text{miss}}$, in the fully inclusive $t\bar{t}$ event sample, for particle-flow reconstruction (solid triangles) and for calorimeter reconstruction (open squares).

The $t\bar{t}$ event sample addresses the case of real missing transverse energy. The missing transverse energy response, defined as the relative average difference between the reconstructed and the true E_T^{miss} is shown in Fig. 17. Although without any a posteriori correction, the response from particle-flow reconstruction is determined to within about $\pm 5\%$ of the true value, as soon as the true E_T^{miss} is above 20 GeV.

For those events containing at least 20 GeV of true E_T^{miss} , the distribution of the difference between the reconstructed and the true E_T^{miss} is displayed in Fig. 18a and also for few representative bins in true E_T^{miss} in Figs. 18b-d. In each true E_T^{miss} bin, a Gaussian fit to this distribution is performed. The relative E_T^{miss} resolution, obtained from this fit and presented in Fig. 20a as a function of the true E_T^{miss} , is improved by almost a factor two with respect to the calorimetric determination, irrespective of the true missing transverse missing energy.

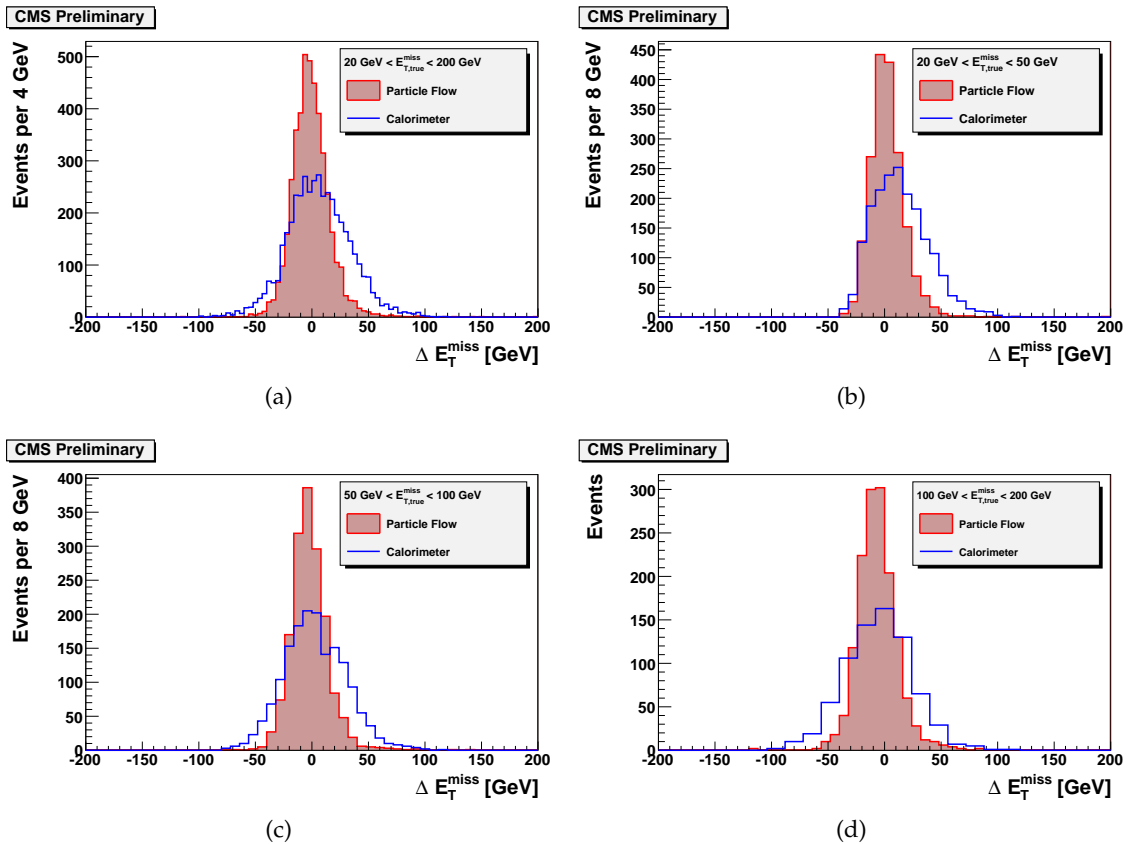


Figure 18: Distributions of $E_{T,\text{reco}}^{\text{miss}} - E_{T,\text{true}}^{\text{miss}}$ for (a) $20 \text{ GeV} < E_{T,\text{true}}^{\text{miss}} < 200 \text{ GeV}$; (b) $20 \text{ GeV} < E_{T,\text{true}}^{\text{miss}} < 50 \text{ GeV}$; (c) $50 \text{ GeV} < E_{T,\text{true}}^{\text{miss}} < 100 \text{ GeV}$; (d) $100 \text{ GeV} < E_{T,\text{true}}^{\text{miss}} < 200 \text{ GeV}$; in the $t\bar{t}$ event sample, for particle-flow reconstruction (solid histograms) and for calorimeter reconstruction (open histograms).

The same kind of improvement is observed for the direction of the missing transverse energy vector, as exemplified in Figs. 19a-d and Fig. 20b.

6 Conclusion

The design of CMS detector is almost ideally suited to attempt particle-flow event reconstruction at LHC: the large magnetic field, the possibility of having large tracking efficiency while keeping a low fake rate, the fine electromagnetic calorimeter granularity and the ability to

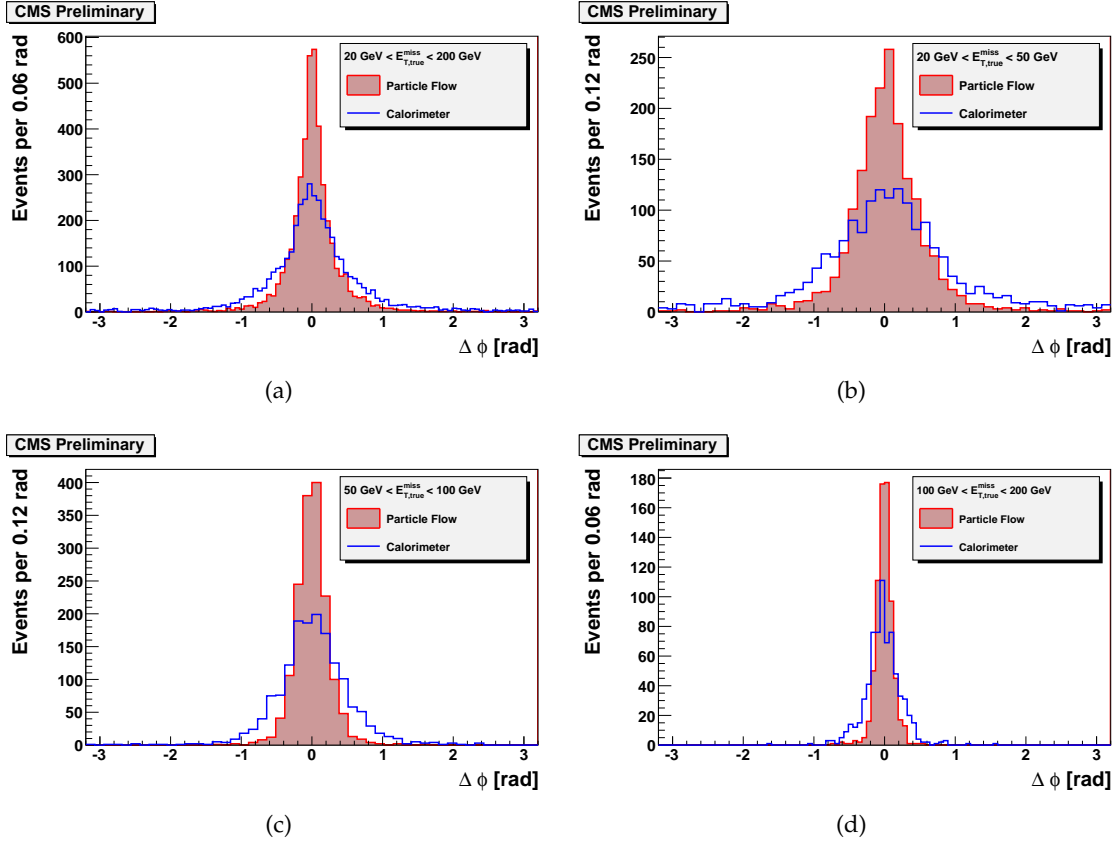


Figure 19: Distributions of $\phi_{\text{reco}} - \phi_{\text{true}}$ for (a) $20 \text{ GeV} < E_{T,\text{true}}^{\text{miss}} < 200 \text{ GeV}$; (b) $20 \text{ GeV} < E_{T,\text{true}}^{\text{miss}} < 50 \text{ GeV}$; (c) $50 \text{ GeV} < E_{T,\text{true}}^{\text{miss}} < 100 \text{ GeV}$; (d) $100 \text{ GeV} < E_{T,\text{true}}^{\text{miss}} < 200 \text{ GeV}$; in the $t\bar{t}$ event sample, for particle-flow reconstruction (solid histograms) and for calorimeter reconstruction (open histograms).

reconstruct all muons with high purity have been shown to be key elements in this respect. When used in combination in a meticulous reconstruction of the full particle content of the events, they lead to an improved expected performance for jets, taus and E_T^{miss} .

Beyond this potential improvement, the combination of the information from all sub-detectors also allows various systematic uncertainties and biases (such as dependence on calorimeter response, noise and resolutions, tracking efficiency or particle content) to be substantially reduced and kept under control. This intrinsic robustness and built-in redundancy will be a major asset when it comes to analyse the first data delivered by the LHC.

References

- [1] CMS Collaboration, “CMS: Detector Performance and Software,” *CERN/LHC 2006-001* (2006).
- [2] CMS Collaboration, “Track Reconstruction in the CMS tracker (in preparation),” *CMS-PAS TRK-09-001* (2009).
- [3] CMS Collaboration, “The CMS experiment at the CERN LHC,” *Journal of Instrumentation* **3** S08004 (2008).

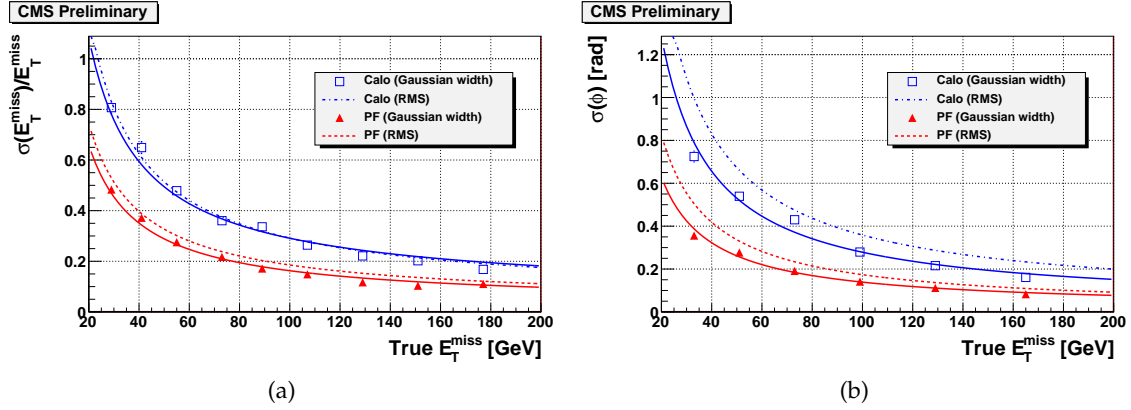


Figure 20: E_T^{miss} reconstruction performance in the $t\bar{t}$ event sample, for particle-flow reconstruction (triangles) and for calorimeter reconstruction (open squares). (a) $\sigma(E_T^{\text{miss}})/E_T^{\text{miss}}$ versus the $E_{T,\text{true}}^{\text{miss}}$ between 20 GeV and 200 GeV. (b) Gaussian fitted ϕ resolution versus the $E_{T,\text{true}}^{\text{miss}}$ of the event between 20 GeV and 200 GeV. A Gaussian fit is performed to each $E_{T,\text{true}}^{\text{miss}}$ bin and the triangles (open squares) represent the sigma of that fit; the solid curves represent a fit through those points. In contrast, the dashed (dash-dotted) curves use the RMS of each $E_{T,\text{true}}^{\text{miss}}$ bin, rather than the sigma of the Gaussian fit, and demonstrate the effect of any non-Gaussian tails.

- [4] **CMS** Collaboration, “Reconstruction of Electron Tracks With the Gaussian-Sum Filter in the CMS tracker at LHC,” *CMS Note RN 2003-001* (2003).
- [5] T. Sjostrand, S. Mrenna, and P. Skands, “PYTHIA 6.4 Physics and Manual,” May, 2006.
- [6] **GEANT4** Collaboration, “Geant 4 - A simulation toolkit,” *Nucl. Instrum. and Methods* **506** (2003).
- [7] **CMS** Collaboration, “Plans for Jet Energy Corrections at CMS,” *CMS-PAS JME-07-002* (2007).
- [8] **CMS** Collaboration, “Performance of Jet Reconstruction with Charged Tracks only,” *CMS-PAS JME-08-001* (2008).
- [9] **CDF** Collaboration, A. Abulencia et al., “Measurement of $\sigma(p\bar{p} \rightarrow Z) \text{Br}(Z \rightarrow \tau^+\tau^-)$ in $p\bar{p}$ collisions at $\sqrt{s}=1.96$ TeV,” *Phys. Rev.* **D75** (2007) 092004.
- [10] C. Gallea, “Measurement of $\sigma(p\bar{p} \rightarrow Z + X) \text{Br}(Z \rightarrow \tau^+\tau^-)$ and search for Higgs bosons decaying to $\tau^+\tau^-$ at $\sqrt{s}=1.96$ TeV”. PhD thesis, NIKHEF, 2008.
- [11] **CMS** Collaboration, “Tau reconstruction and identification with particle-flow techniques using the CMS detector at LHC,” *CMS-PAS PFT-08-001* (2008).
- [12] **CMS** Collaboration, “Tau Jet Reconstruction and Tagging in CMS,” *Eur. Phys. J. C direct Electronic Only* (2006).
- [13] **CMS** Collaboration, “Missing ET Performance,” *CMS-PAS JME-07-001* (2009).