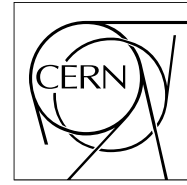


The Compact Muon Solenoid Experiment

Analysis Note

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



30 April 2009 (v3, 04 May 2009)

Jet Reconstruction Performance at CMS

V. Chetluru

Fermilab, Batavia, IL, USA

F. Pandolfi

“Sapienza” Università di Roma and INFN Sezione di Roma, Rome, Italy

P. Schieferdecker

Institut für Experimentelle Kernphysik, Universität Karlsruhe (TH), Karlsruhe, Germany

M. Zielinski

University of Rochester, Rochester, NY, USA

Abstract

CMS employs three different approaches to reconstruct jets originating from quarks and gluons in the final state: calorimeter jet reconstruction based only on the energy deposits in the calorimeters, the “Jet-Plus-Track” algorithm, which improves the energy measurement of calorimeter jets by exploiting the kinematics of tracks associated to them, and the “Particle Flow” reconstruction, which attempts to reconstruct and calibrate each particle arising from LHC collisions individually combining the information from all sub-detectors prior to the formation of jets. The performance of the various combinations of these techniques with different jet clustering algorithms and jet size parameters is evaluated and compared for simulated QCD dijet events from the “Summer08” production. Two jet clustering algorithms not previously studied at CMS are presented for the first time: the Anti- k_T and Cambridge/Aachen algorithms. Both are known to be infrared- and collinear-safe, and especially the Anti- k_T algorithm is found to perform very well and show strikingly similar behavior compared to the theoretically disfavored Iterative Cone algorithm. We therefore propose for CMS to adopt the Anti- k_T algorithm as part of the official reconstruction chain, and to eventually consider it as a replacement of the Iterative Cone algorithm after further studies.

1 Introduction

Quarks and gluons arise in the majority of final states to be studied at CMS, but can not be observed directly: they fragment into stable hadrons first, which then produce signals in the electromagnetic- (ECAL) and hadronic- (HCAL) calorimeters and, if they are charged, the pixel- and silicon-strip tracking detectors. These signals are subsequently clustered into collimated objects composed of stable particles, called “jets”, which are expected to yield a good representation of both the parton-level and the hadron showers emerging from the hard interaction. This note describes the different reconstruction techniques and jet clustering algorithms employed at CMS, and evaluates and compares the performance of the various combinations.

As the much anticipated Particle-Flow [1] reconstruction is available for the first time to form jets, a significant fraction of this document is devoted to study their properties and compare them to other types of jets. Note however that the algorithm is still under development, and has in fact received numerous updates and improvements available only in more recent CMS software versions at the time of writing. We will therefore update results presented in this note when the jet samples processed with the latest CMS reconstruction software become available and more representative of the performance at LHC startup.

Section 2 details the simulation samples used for the studies presented in this note. The different reconstruction techniques and jet clustering algorithms available to CMS are described in Section 3. Section 4 takes a closer look at the composition of Particle-Flow jets, benchmarking the ability of the algorithm to correctly estimate the multiplicity, kinematics, and particle-type of jet constituents. The jet p_T response, jet energy- and jet position-resolution are studied and compared among the various reconstruction techniques and clustering algorithms in detail in Section 5 using simulated QCD dijet events.

2 Samples

The MC samples used are from the “Summer08” production: for an assumed centre-of-mass energy for the pp collisions of $\sqrt{s} = 10$ TeV, they were simulated with the CMSSW_2_1_X and subsequently reconstructed with the CMSSW_2_2_X series of the CMS software. The resulting QCD dijet sample consist of 1M events generated with PYTHIA [2] divided into 21 \hat{p}_T bins, listed in Table 1.

DBS dataset name	σ [pb]	events
/QCDDiJetPt0to15/Summer08_IDEAL_V11_redigi_v3/GEN-SIM-RECO	5.16e+10	110340
/QCDDiJetPt15to20/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO	9.49e+08	96120
/QCDDiJetPt20to30/Summer08_IDEAL_V11_redigi_v2/GEN-SIM-RECO	4.01e+08	108720
/QCDDiJetPt30to50/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO	9.47e+07	111600
/QCDDiJetPt50to80/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO	1.22e+07	97875
/QCDDiJetPt80to120/Summer08_IDEAL_V11_redigi_v2/GEN-SIM-RECO	1.62e+06	48600
/QCDDiJetPt120to170/Summer08_IDEAL_V11_redigi_v2/GEN-SIM-RECO	2.56e+05	54135
/QCDDiJetPt170to230/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO	4.83e+04	48600
/QCDDiJetPt230to300/Summer08_IDEAL_V11_redigi_v2/GEN-SIM-RECO	1.06e+04	59400
/QCDDiJetPt300to380/Summer08_IDEAL_V11_redigi_v3/GEN-SIM-RECO	2.63e+03	68688
/QCDDiJetPt380to470/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO	7.22e+02	46656
/QCDDiJetPt470to600/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO	2.41e+02	18144
/QCDDiJetPt600to800/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO	6.25e+01	20196
/QCDDiJetPt800to1000/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO	9.42e+00	16920
/QCDDiJetPt1000to14000/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO	2.34e+00	34804
/QCDDiJetPt1400to1800/Summer08_IDEAL_V11_redigi_v2/GEN-SIM-RECO	1.57e-01	34804
/QCDDiJetPt1800to2200/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO	1.38e-02	29104
/QCDDiJetPt2200to2600/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO	1.30e-03	22320
/QCDDiJetPt2600to3000/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO	1.14e-04	17040
/QCDDiJetPt3000to3500/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO	8.43e-06	21360
/QCDDiJetPt3500toInf/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO	1.81e-07	21360

Table 1: PYTHIA QCD dijet samples from the “Summer08” production, divided into 21 \hat{p}_T bins.

3 Jet Reconstruction at CMS

Three different types of jets are reconstructed at CMS, based on the information of which sub-detectors is used and how the individual contributions are combined to form the inputs to the jet finding: calorimeter jets, Jet-Plus-Track (JPT) jets, and Particle-Flow (PFlow or PF) jets. All three are briefly introduced in this Section, followed by descriptions of all jet clustering algorithms currently available at CMS. To evaluate their performance, “generator jets” are reconstructed as well by applying each clustering algorithm to all stable particles from the MC truth record. These generator jets are then associated to jets reconstructed from the simulated detector signals, based on the spatial separation between the two jet axes in $\eta - \phi$ -space by requiring

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3, \quad (1)$$

such that no two generator jets are paired with the same reconstructed jet and vice versa. Since generator jets are based on stable MC *particles* after hadronization, reconstructed jet energies corrected to correspond on average to these generator jet energies are considered as corrected to the “particle-level” or “hadron-level”.

3.1 Calorimeter Jet Reconstruction

Calorimeter jets are reconstructed using energy deposits in the electromagnetic and hadronic calorimeter cells, combined into *calorimeter towers* as inputs: a calorimeter tower consists of one or more HCAL cells and the geometrically corresponding ECAL crystals. In the barrel region of the calorimeters ($|\eta| < 1.4$), the unweighted sum of one single HCAL cell and 5x5 ECAL crystals form a projective calorimeter tower. The association between HCAL cells and ECAL crystals is slightly more complex in the endcap regions of the electromagnetic calorimeter ($1.4 < |\eta| < 3.0$). Beyond the coverage of the ECAL, each calorimeter tower corresponds to one hadronic calorimeter cell. To reject electronic noise, the deposit in a given cell is only added to the tower energy if it passes the “Scheme B” energy thresholds [3], which are listed in Table 2 for the various detector regions. Towers are only considered for further clustering into jets if they furthermore fulfill $E_T > 0.5$ GeV to reduce the impact of additional Pile-Up (PU) interactions in the event. Both requirements will be revisited and optimized soon once the hardware settings of the CMS calorimeters for first collision data are finalized [4].

Scheme	HB [GeV]	HO [GeV]	HE [GeV]	$\sum EB$ [GeV]	$\sum EE$ [GeV]
B	0.90	1.10	1.40	0.20	0.45

Table 2: Energy thresholds (in GeV) for calorimeter noise suppression “Scheme B”. $\sum EB$ and $\sum EE$ refer to the sum of ECAL energy deposits associated with the same tower in the barrel and endcap respectively.

3.2 Calorimeter Jet-Plus-Tracks Reconstruction

The *Jet-Plus-Tracks* (JPT) algorithm [5] exploits the excellent performance of the CMS tracking detectors to improve the p_T response and resolution of calorimeter jets. The algorithm is currently limited to fixed-cone-type jet finders like the Iterative Cone algorithm described below, but this limitation can in principle be overcome in future updates. Calorimeter jets are reconstructed first as described in Section 3.1, then charged particle tracks are associated with each jet based on spatial separation in η - ϕ between the jet axis and the track momentum measured at the interaction vertex. The associated tracks are projected onto the surface of the calorimeter and classified as *in-cone tracks* if they point to within the jet cone around the jet axis on the calorimeter surface. If the 3.8 T magnetic field of CMS has instead bent the track out of the jet cone projection on the calorimeter surface, it is classified as a *out-cone track*. The momenta of both in-cone and out-cone tracks are then added to the energy of the associated calorimeter jet. For in-cone tracks the expected average energy deposition in the calorimeters is furthermore subtracted based on the momentum of the track and the hypothesis that it originates from a charged pion. The corresponding map describing average calorimeter energy response of charged pions as a function of p_T and η is derived from di-pion particle gun MC samples and will be accessible in real data through events with isolated charged (hadron) tracks. Note that the direction of the axis of the original calorimeter jet is not corrected by the algorithm in its current implementation.

3.3 Particle Flow Jet Reconstruction

The *Particle Flow* (PF) algorithm [1] aims to reconstruct, identify and calibrate each single particle in the event individually by combining all CMS sub-systems. *Particles* are reconstructed as a combination of charged particle

tracks and clusters in the electromagnetic and hadronic calorimeters, as well as signals in either of the two CMS pre-shower detectors and the muon system. Depending on which of the detector systems contribute to a single particle, it is identified as either an electron (track+ECAL), muon (track+ECAL+HCAL+Muon System), photon (ECAL), charged hadron (track+ECAL+HCAL), or neutral hadron (HCAL). The algorithm employs sophisticated strategies to handle ambiguities stemming from overlapping detector signals, to make sure that no measurement is counted twice. Based on the particle type, the energy of each particle is calibrated, where hadrons are assumed to be pions. Both the calibrations of charged and neutral hadrons are currently based on single pion MC samples and determined through fits based on the real hadron energy provided by the simulation. The calibration of charged hadrons, which on average contribute about 60 % of the jet energy, can in principle be derived from real collision data, selecting events which contain isolated charged particle tracks. Neutral hadrons can only be calibrated using simulation data, but mis-calibration impacts the overall jet reconstruction only mildly, as undecayed neutrals only contribute about 10 % [1] of the jet energy on average (see Section 4 for a more in-depth analysis of jet composition). The goal of the PF reconstruction is to produce fully calibrated inputs to the jet reconstruction, such that the resulting jets require little or no further a-posteriori energy corrections. Like for JPT jets, the p_T resolution is expected to be improved w.r.t. calorimeter jets mostly through the use of the tracking detectors, which are expected to lead as well to better jet position (η , ϕ) resolution for PF jets. The very good granularity of the CMS electromagnetic calorimeters is furthermore exploited in a more optimal way than for the other two reconstruction techniques.

3.4 Jet Clustering Algorithms

To carry out accurate comparisons between parton-level predictions from theory and hadron-level observations in collider experiments, well defined jet finding procedures are required. As jet-finding is an approximate attempt to reverse-engineer the quantum mechanical processes of fragmentation and hadronization, it is not a unique procedure. Hence several different jet finders, also referred to as jet algorithms were proposed over the last decades in the context of collider event reconstruction. Two distinct types of jet finders are most commonly used in the high energy physics community, *cone-type* and *sequential clustering* algorithms.

Three sequential clustering algorithms are available at CMS to reconstruct jets: k_T [6], Cambridge/Aachen [7], and Anti- k_T [8]. They are all based on successive pair-wise recombination of particles (input four vectors) according to the distance between any two particles i and j , d_{ij} and the distance of any particle i to the beam, d_{iB} , which are defined as

$$d_{ij} = \min \left(k_{Ti}^{2p}, k_{Tj}^{2p} \right) \frac{\Delta_{ij}^2}{D^2}, \quad (2)$$

$$d_{iB} = k_{Ti}^{2p}, \quad (3)$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and k_{Ti} , y_i , and ϕ_i are respectively the transverse momentum, rapidity, and azimuth of particle i . Each clustering algorithm identifies the smallest distance among all d_{ij} and d_{iB} in the list of input particles, and if it is a d_{ij} recombines particles i and j to form one single new particle by adding their four momenta. If the smallest distance is a d_{iB} , particle i is removed from the list of particles and called a jet. After each step all distances are recalculated and the procedure is repeated until no particles are left to be clustered. The jet size parameter D rescales the distances d_{ij} w.r.t. the d_{iB} such that any pair of final jets a and b are separated by at least $\Delta_{ab}^2 = D^2$. According to the general definition of the clustering metrics in Equations 2 and 3 the parameter p , which governs the relative power of energy versus geometrical scales, distinguishes between the three jet finders: $p = 1$ yields the k_T algorithm, while $p = 0$ and $p = -1$ correspond to the Cambridge/Aachen and Anti- k_T algorithms respectively. All three algorithms are infrared- and collinear safe: neither soft emissions nor collinear splitting change the result of the clustering, leading to a robust event interpretation in terms of partons and allowing the application of the algorithms in theoretical calculations for comparisons with experimental data. Despite their appealing simplicity and long list of advantages, sequential clustering algorithms are not widely used at current hadron collider experiments, historically due to prohibitively growing CPU requirements for large numbers of input particles to be clustered per event. The CMS software however interfaces to the FASTJET [9, 10] package, which provides novel implementations of these algorithms. Yielding the exact same results at dramatically reduced processing times, they become the fastest available jet finders suitable to efficiently reconstruct the large-multiplicity events to be delivered by the LHC.

The CMS jet group supports also two cone-type algorithms, Iterative Cone [11] and SIScone [12]. The Iterative Cone algorithm uses the most energetic particle as a seed and clusters all particles within a cone of radius R ¹⁾

¹⁾ R is the jet size parameter for cone jets, corresponding to the D parameter for sequential clustering algorithms introduced in Equation 2

around the seed axis into a trial jet. The sum of all constituent particles of the trial jet is formed, and the corresponding axis is compared to the seed. If the two axes agree within precision, the cone is considered *stable*, the trial jet is called a jet, its constituents are removed from the list of particles, and the algorithm proceeds with the most energetic remaining particle as the next seed. Otherwise, the procedure is repeated iteratively with the axis of the trial jet as the new seed until it converges to yield a stable cone. The algorithm proceeds until no seed with energy above 1 GeV remains. While this algorithm is simple, features fast execution and yields good p_T resolution, it is neither collinear- nor infrared-safe, like most commonly used cone-type algorithms. Therefore the “Seedless Infrared-Safe Cone” jet algorithm (SISCone [12], also provided via FASTJET) was proposed, which employs an exhaustive non-iterative approach which provably finds all stable cones, is infrared- and collinear-safe, and features moderate computational execution time.

CMS currently employs Iterative Cone with $R = 0.5$, SISCone with $R = 0.5$ and $R = 0.7$, and k_T with $D = 0.4$ and $D = 0.6$: they are part of the default CMS reconstruction chain, and jet energy corrections are regularly provided by the JetMET group. The Anti- k_T and Cambridge/Aachen jet finders have only recently been interfaced to the CMS software, and their performance is studied and compared to the available jet finders here for the first time in CMS. Special interest is devoted to the Anti- k_T algorithm for it is expected [8] to have similar performance and features as the Iterative Cone algorithm, while also being infrared- and collinear-safe.

3.5 Jet Energy Corrections

Jets are corrected to the *particle-level* at CMS following a factorized scheme [13], where three different corrections are applied successively to each jet: the *offset correction* [14] to subtract energy originating from electronic noise and PU, the *relative correction* [15] to equalize the response of all jets as a function of jet η w.r.t. to the average response in the calorimeter barrel, and finally the *absolute correction* [16] to account for the difference of the response of a reconstructed jet in the barrel and the associated generator jet. Data-driven strategies to derive each individual correction directly from collision data were successfully employed at the Tevatron and have been proposed, studied, and validated for CMS as well. For the performance studies in this note however, the relative and absolute corrections derived from generator jets in QCD dijet MC samples and officially released for the Summer08 MC production are used.

4 Composition of Particle Flow Jets

In this section we take a closer look at the ability of the Particle-Flow algorithm to reconstruct the composition of jets, in terms of the jet constituent (particle) multiplicities, kinematics, and types. Note that PF jets in this section are matched to generator jets within $\Delta R < 0.1$ to account for their improved position resolution compared to calorimeter jets, see Section 5.

Figure 1 compares the energy contributions to PF jets and generator jets from the three dominant sources: About 60 % of the jet energy arises from charged hadrons, which is slightly overestimated by the PFlow algorithm by about 10 %. The structure as a function of η in the reconstructed PF jet spectrum follows the material budget of the CMS tracking system, and is therefore attributed to photon conversions into e^+e^- pairs, which are not yet treated specifically and likely to be reconstructed as charged hadrons. The fact that the fraction of photons is overestimated (≈ 30 % compared to ≈ 20 % at generator level) and the neutral hadron fraction is underestimated accordingly (≈ 10 % compared to ≈ 20 % at generator level) is a direct consequence of a conscious decision within the PF reconstruction: in the case of excess energy deposits in the calorimeter w.r.t. the track measurement linked to the cluster(s), precedence is given to the creation of a photon over a neutral hadron. In addition, the ≈ 20 % neutral hadron fraction at generator level is effectively reduced at detector level, as ≈ 8 % originate from K_S^0 and Λ^0 particles: both decay according to their respective mean free path lengths of about 10 cm and 20 cm well within the active tracking volume, and predominantly into charged hadrons²⁾ and π^0 s (photons). The shoulders in the photon energy contributions for PF jets at the ECAL endcap boundaries ($|\eta| \approx 3.0$) are attributed to the yet to be included calibration of the pre-shower detectors. Tracker- and ECAL- coverages end at $|\eta| > 2.6$ and $|\eta| > 3.0$ respectively, hence PF particles (clusters) beyond are identical to the HCAL calorimeter towers from the standard reconstruction. Electrons and muons arise only in a marginal number of jets in the QCD sample under study and their contribution is therefore negligible.

²⁾ The CMS track reconstruction in future software versions can efficiently reconstruct tracks which do not originate from the interaction vertex.

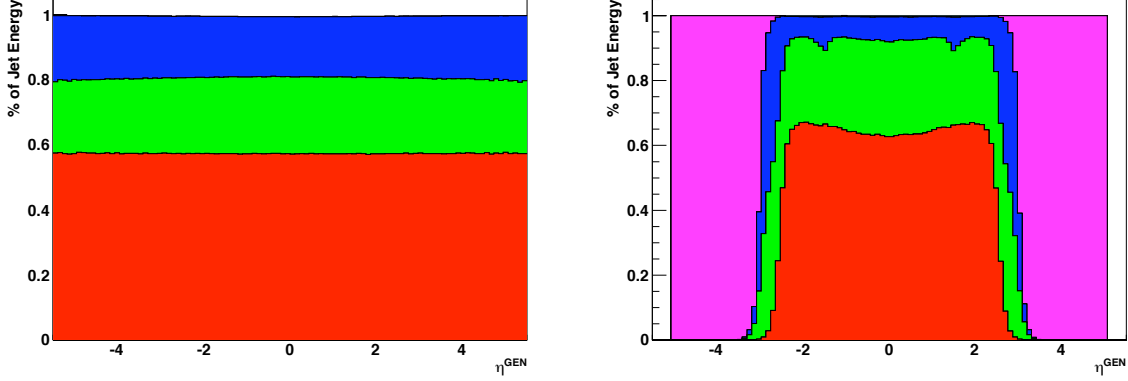


Figure 1: Average energy composition of generator jets (*left*) and PF jets (*right*) as a function of the pseudo-rapidity η^{gen} for jets in the range $50 < p_T^{\text{gen}} < 80$ GeV. The various contributions correspond (from *top to bottom*) to **neutral hadrons** (blue), **photons** (green), and **charged hadrons** (red). Particles with $|\eta| > 3.0$ are beyond the coverage of both the tracker and the ECAL and therefore directly correspond to reconstructed **HCAL clusters** (magenta).

The average particle multiplicity and the average particle p_T per jet are shown in Figure 2 for both generator- and PF-jets as a function of p_T^{gen} in the barrel ($0 < |\eta| \leq 1.4$). With rising jet p_T , jets tend to get narrower and the particle density rises, such that the particle reconstruction efficiency decreases due to the finite granularity of the tracking system and the ECAL. The PFlow algorithm compensates this effect however by reconstructing particles with on average larger transverse momenta, as charged particles which are not reconstructed as tracks lead to excess energies in the calorimeter, which in turn are interpreted as additional particles themselves. Thus, the PFlow algorithm achieves the desirable feature to reconstruct jets based dominantly on tracking at low p_T , but based dominantly on calorimetry at high p_T .

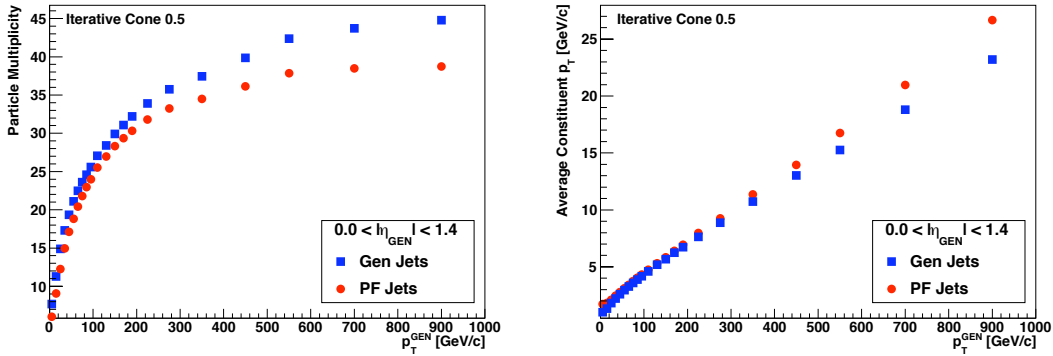


Figure 2: Average particle multiplicity (*left*) and average particle p_T (*right*) for generator- (squares) and PF-jets (circles) as a function of p_T^{gen} .

As demonstrated by Figure 2, even high p_T jets are typically comprised of many individual constituents with on average relatively small transverse momenta. If these are charged they can be measured with high efficiency and low fake-rate with the tracking system very accurately, and pre-calibrated under the charged pion hypothesis prior to feeding them to the jet reconstruction. It can therefore be expected that the response as defined in Equation 4 below is systematically biased by the *charged fraction* R_{ch} for PF jets, which is defined as the fraction of the jet energy contributed by charged particles. This impact of R_{ch} on the jet response is demonstrated in Figure 3 both for the barrel and the endcaps. The R_{ch} distributions for both generator jets and PF jets in the barrel and endcap regions are shown in Figure 4. Note that even PF jets with $R_{\text{ch}} = 0$ for $2.6 < |\eta| < 3.0$ show slightly higher response than uncorrected calorimeter jets due to the charged hadron calibration performed in the PFlow algorithm. The observed behavior in the barrel and endcaps strongly suggests that the resolution of PFlow jets can be further improved by a dedicated energy correction which takes the charged fraction R_{ch} into account [17].

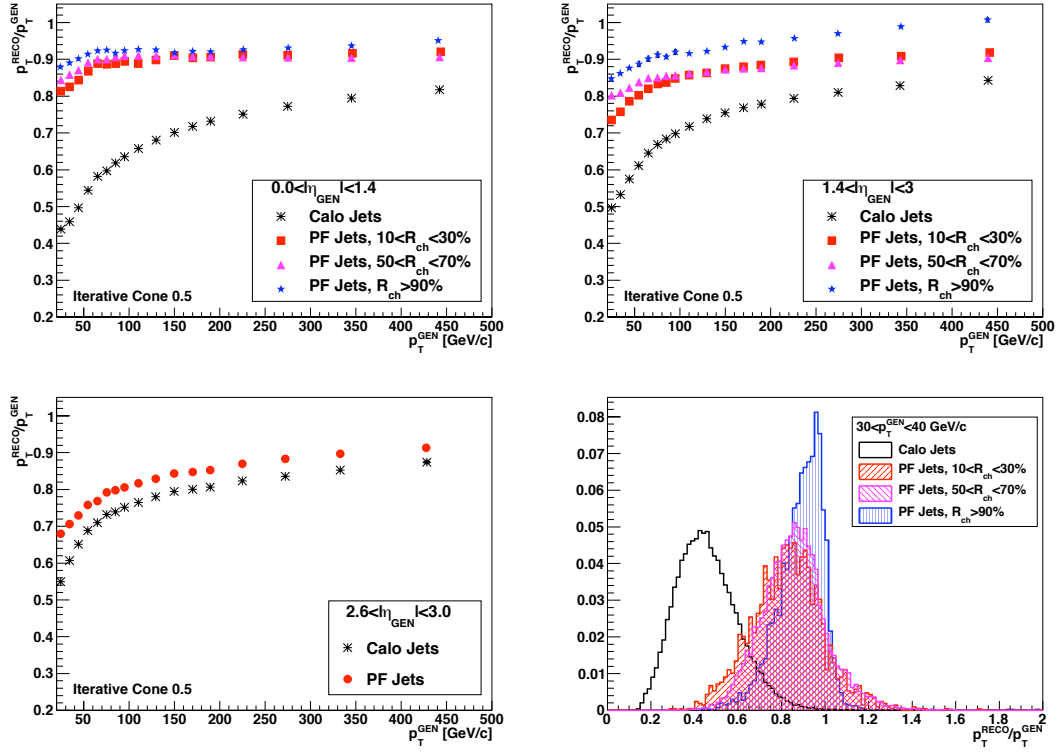


Figure 3: Response for PF jets with different charged fractions R_{ch} in the barrel (*top left*) and endcaps (*top right*) as a function of p_T , compared to uncorrected calorimeter jets. The *bottom left* plot shows the response for $2.6 < |\eta| < 3.0$, where PF jets show slightly higher response than calorimeter jets despite $R_{\text{ch}} = 0$. The jet response distributions for PF jets with different R_{ch} as well as calorimeter jets are shown in the *bottom right* plot for $30 < p_T < 40$ GeV in the barrel.

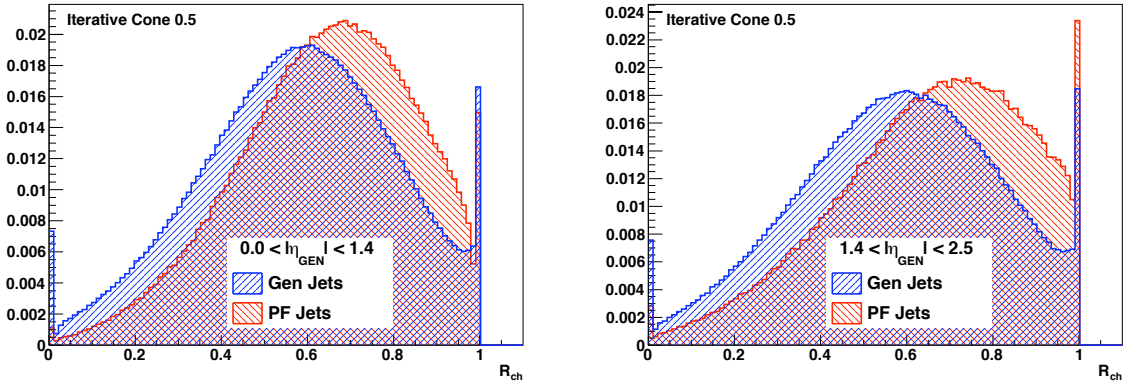


Figure 4: The fraction of generator and PF jet energy contributed by charged particles, R_{ch} , for the barrel ($0 < |\eta| \leq 1.4$, *left*) and endcaps (within tracker coverage: $1.4 < |\eta| < 2.5$, *right*).

5 Jet Reconstruction Performance in QCD Dijet Events

The performance of the various jet reconstruction types and algorithms is evaluated in QCD dijet events by comparing the jet p_T response and jet p_T resolution, as well as the jet position resolutions in η and ϕ . Sections 5.1 and 5.2 are dedicated to the three algorithms currently used by CMS (Iterative Cone, SISCone, and k_T), while Section 5.3 addresses the performance of the Anti- k_T and Cambridge/Aachen jet finders applied to CMS simulation events for the first time.

5.1 Energy Response and Resolution

The (relative) jet p_T response is defined for any pair of spatially matched generator and reconstructed jets as

$$p_T/p_T^{\text{gen}}, \quad (4)$$

where p_T refers the raw or jet energy corrected transverse momentum of the reconstructed jet. In order to derive the response as a function of p_T^{gen} , several p_T^{gen} bins are evaluated, and either the *mean* of the distribution itself or of a Gaussian fit to it is interpreted as the response associated with $\langle p_T^{\text{gen}} \rangle$ in the corresponding bin. The response as a function of any other variable like η or ϕ is defined accordingly. Figure 5 shows the response for calorimeter-, PF-, and JPT-jets, each reconstructed with a different jet algorithm to illustrate more than one of them, in the bin $35 < p_T^{\text{gen}} < 45$ GeV. Note that only Iterative Cone $R = 0.5$ jets are currently available for the JPT reconstruction. The Gaussian fits to the $\pm 5 \cdot \text{RMS}$ range around the *mean* are shown as well and are found to yield a good description of the data points, except for PF jets. Several reasons for slightly more pronounced tails and non-Gaussian features in the core of PF response distributions were identified and taken into account in more recent versions of the PFlow algorithm, most notably a more stringent rejection of fake tracks and increased track reconstruction efficiency, respectively.

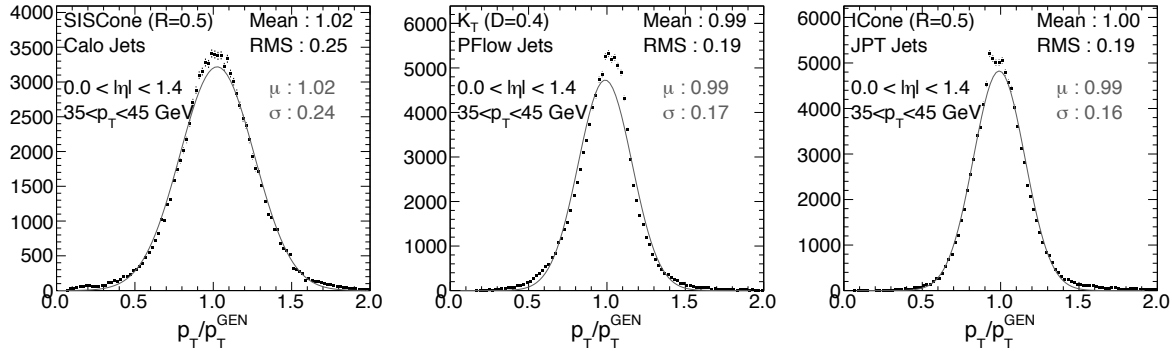


Figure 5: Jet p_T response for calorimeter jets (*left*), PF jets (*middle*) and JPT jets (*right*) in the range $35 < p_T^{\text{gen}} < 45$ GeV and different jet algorithms and jet sizes as indicated. The Gaussian fits to the $\pm 5 \cdot \text{RMS}$ range around the *mean* of the distributions show reasonable agreement, except for PF jets.

The jet p_T response as a function of p_T^{gen} is shown in Figure 6 for calorimeter jets, PF jets, and JPT jets. All three types of jets yield good correspondence to average response of 1.0 after the application of jet energy corrections. The raw response is also shown and demonstrates the degree of non-linearity as well as the size of the associated energy correction, which is merely residual for both PF and JPT jets, but substantial for calorimeter jets, as expected. The jet response as a function of η is shown in Figure 7 before and after the application of energy corrections. The most non-uniformities are again observed for calorimeter jets, and are in all cases properly addressed by the corrections.

Figure 8 shows the jet p_T resolution as a function of p_T^{gen} for all three types of jets. Two measures of the resolution are compared in each plot: the response as derived from the σ of a Gaussian fit to the $\pm 5 \cdot \text{RMS}$ range around the *mean* of the response distribution in each individual p_T^{gen} bin, and from the *RMS* of it. The result is divided by the μ of the Gaussian fit to the response or by the *mean* of it, for resolutions derived from σ and *RMS* respectively. Reasonably good agreement between the two indicates that these Gaussian fits yield a good measure of the resolution, hence we choose the parameters of the fit to quantify jet p_T resolutions throughout this note.

The relative jet p_T resolutions as a function of p_T^{gen} are fitted to the sum of a constant term c , a stochastic term b , and a noise term a for all combinations of jet type, clustering algorithm, and size parameters considered in this

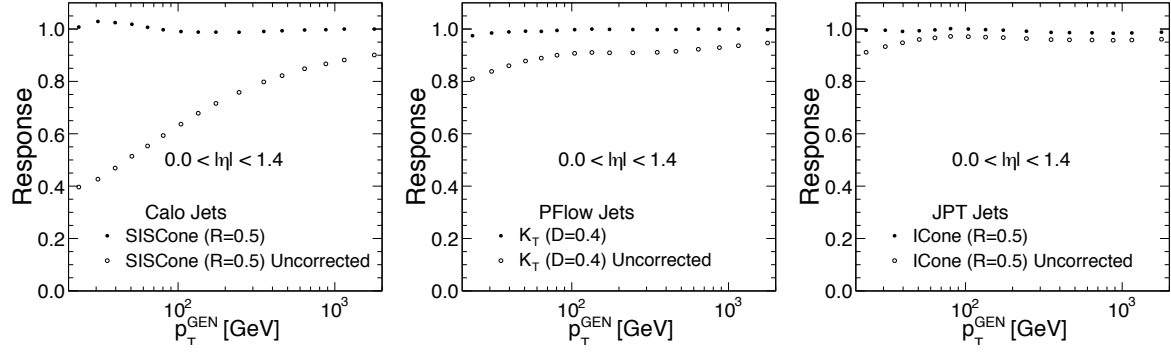


Figure 6: Jet p_T response as a function of p_T^{gen} before (open circles) and after (full circles) application of jet energy corrections, for calorimeter jets (*left*), PF jets (*middle*), and JPT jets (*right*). Different jet algorithms and jet sizes are used as indicated on the plots.

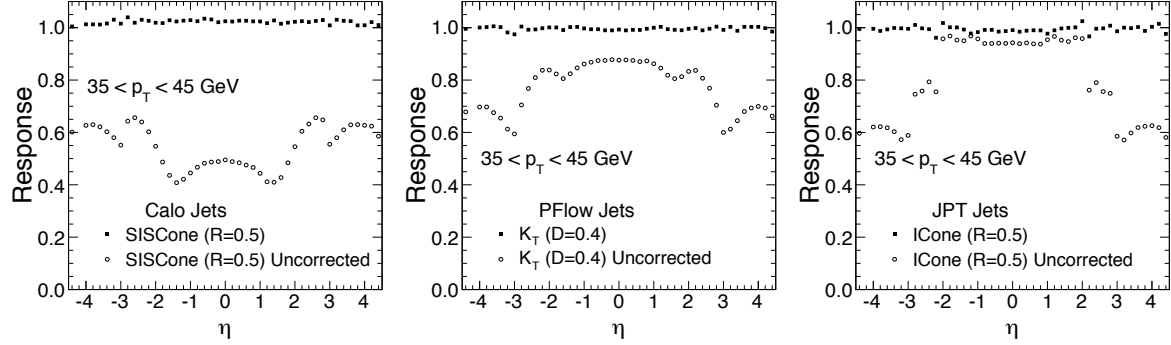


Figure 7: Jet p_T response as a function of η before (open circles) and after (full circles) application of jet energy corrections, for calorimeter jets (*left*), PF jets (*middle*), and JPT jets (*right*). Different jet algorithms and jet sizes are used as indicated on the plots.

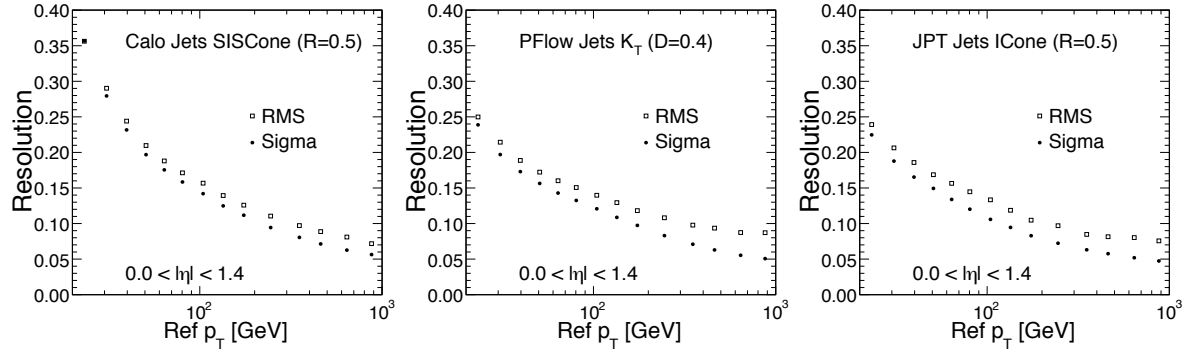


Figure 8: Jet p_T resolution based on the width of a Gaussian fit to (full circles) and the RMS of (open squares) the individual response distributions, for calorimeter jets (*left*), PF jets (*middle*), and JPT jets (*right*). Different jet algorithms and jet sizes are used as indicated on the plots.

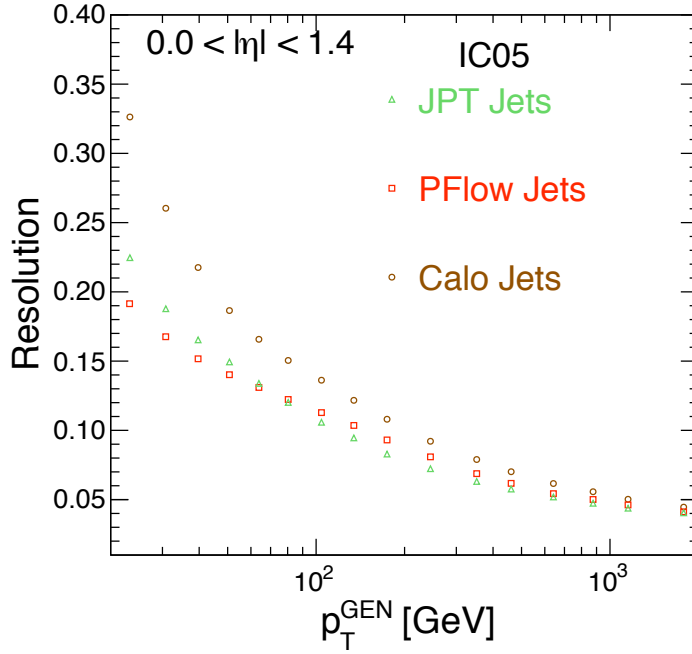


Figure 9: Comparison of jet p_T resolutions for calorimeter jets (circles), PF jets (squares), and JPT jets (triangles), reconstructed with the Iterative Cone ($R = 0.5$) algorithm in the barrel region.

note,

$$\frac{\sigma(p_T/p_T^{\text{gen}})}{\langle p_T/p_T^{\text{gen}} \rangle}(p_T^{\text{gen}}) = \sqrt{\left(\frac{a}{p_T^{\text{gen}}}\right)^2 + \left(\frac{b}{\sqrt{p_T^{\text{gen}}}}\right)^2 + c^2}. \quad (5)$$

The corresponding fits comparing the resolutions among four different detector regions for all three types of jets are shown together with the individual data points in Figure 10. The classification of these regions naturally follow the structure of the CMS detector and its sub-systems: the *barrel*, $0 < |\eta| \leq 1.4$, is covered by the central ECAL and HCAL systems, namely EB and HB; The *endcap* systems of both detectors, EE and HE, reach out to about $|\eta| < 3.0$, but as the tracker coverage ends at approximately $|\eta| = 2.6$, we consider only $1.4 < |\eta| \leq 2.6$, and define the *transition* region between endcaps and *forward* ($3.2 < |\eta| \leq 4.7$) as $2.6 < |\eta| \leq 3.2$. This segmentation appears to be justified according to Figure 10, which shows different resolution for each of these regions for all three types of jets. Note that calorimeter jets in the transition region are significantly impacted as well despite not being affected by tracker coverage: as a jet is classified by the position of its axis, its constituents might effectively extend it into the neighboring region(s).

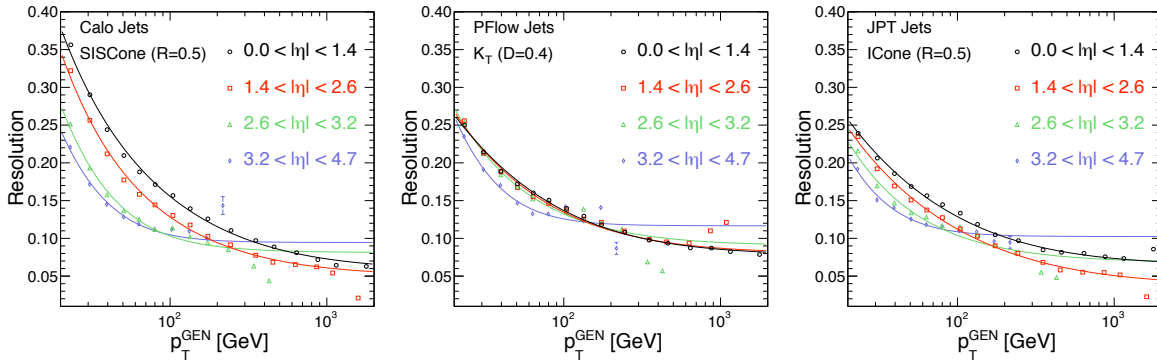


Figure 10: Jet p_T resolution for calorimeter jets (left), PF jets (middle), and JPT jets (right), compared in four different regions of the calorimeters: barrel ($0 < |\eta| \leq 1.4$), endcaps ($1.4 < |\eta| \leq 2.6$), the transition region between endcaps and forward ($2.6 < |\eta| \leq 3.2$), and forward ($3.2 < |\eta| \leq 4.7$). The jet algorithms and sizes used are indicated on the plots.

255

256 The jet p_T resolution for calorimeter jets and PF jets is compared for different clustering algorithms and jet size

parameters and all four detector regions separately in Figures 11 and 12, respectively. The data points are fitted according to the same procedure as described above and illustrated in Figure 10, but the graphical representations of the fits are omitted here. The values of all resolution fits as a function of p_T^{gen} according to Equation 5 however are listed in Tables 3, 4, and 5 for calorimeter-, PF-, and JPT-jets respectively, including the statistical uncertainties of the fit parameters a , b , and c . A description of a method to use these parameters to unsmear a reconstructed jet spectrum under the hypothesis of a specific underlying physics process can be found in [18].

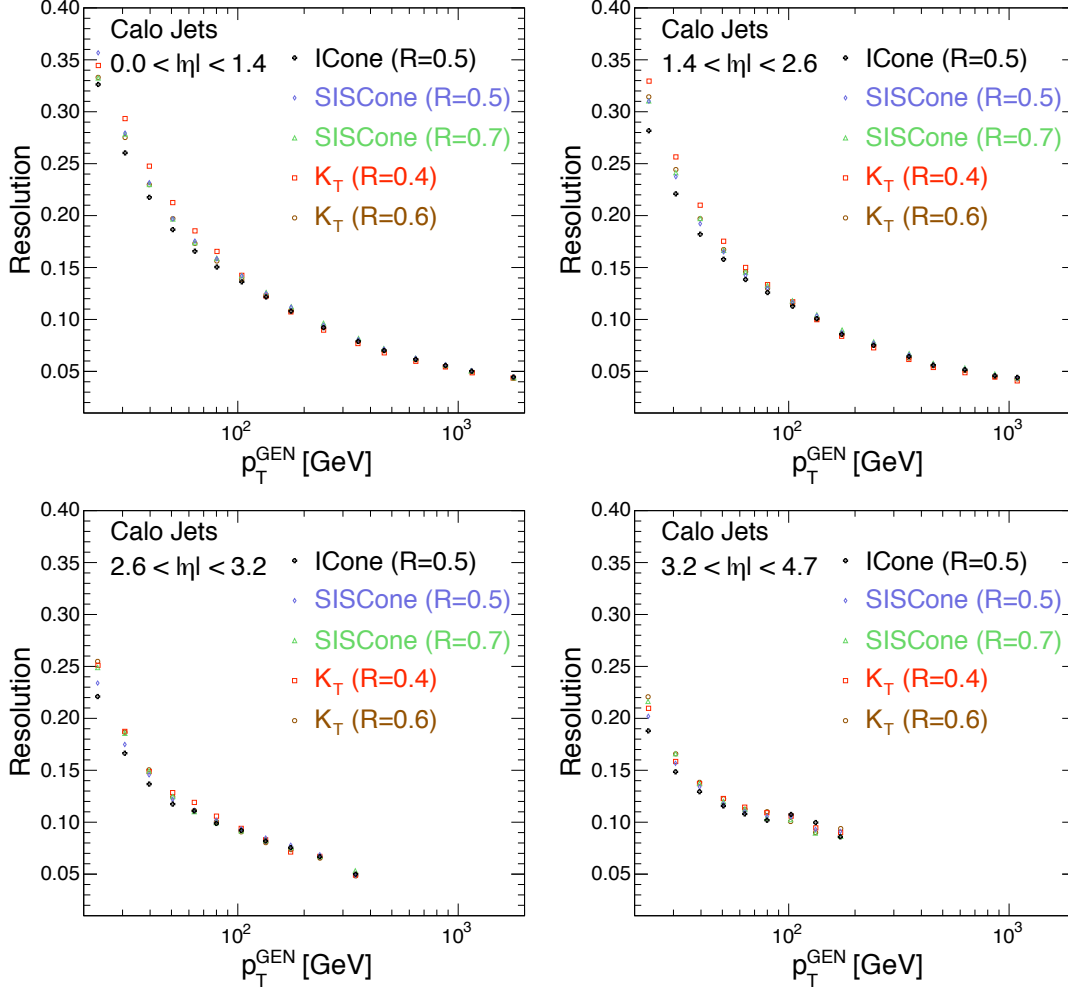


Figure 11: Jet p_T resolution for **calorimeter jets** reconstructed with different jet algorithms and sizes in the barrel (*top left*), endcaps (*top right*), the transition region between endcaps and forward (*bottom left*), and forward (*bottom right*).

5.2 Jet Position Resolution

The jet η and ϕ resolutions are shown in Figures 13 and 14 respectively for calorimeter and PF jets, for different clustering algorithms and jet size parameters, and separately for each detector region. As the current version of the JPT algorithm does not use the additional information from the associated tracks to correct the direction of a jet, the position resolution for JPT jets is identical to calorimeter jets. The data points are derived from $\Delta\eta$ and $\Delta\phi$ distributions in various p_T^{gen} bins, where

$$\Delta\eta = |\eta| - |\eta^{\text{gen}}|, \quad \Delta\phi = \phi - \phi^{\text{gen}}. \quad (6)$$

Because these distributions are fairly symmetric but have substantial non-Gaussian tails, the resolutions are quantified by the *RMS*. Note that η refers to *physics* η : the axis of each jet is calculated w.r.t. the reconstructed primary interaction vertex, not the middle of the detector (0., 0., 0.).

In the barrel and endcap regions, PF jets are found to yield a sizable improvement in both η - and ϕ -resolution, especially at low p_T^{gen} . This is illustrated for the barrel by Figure 15, which explicitly compares the position

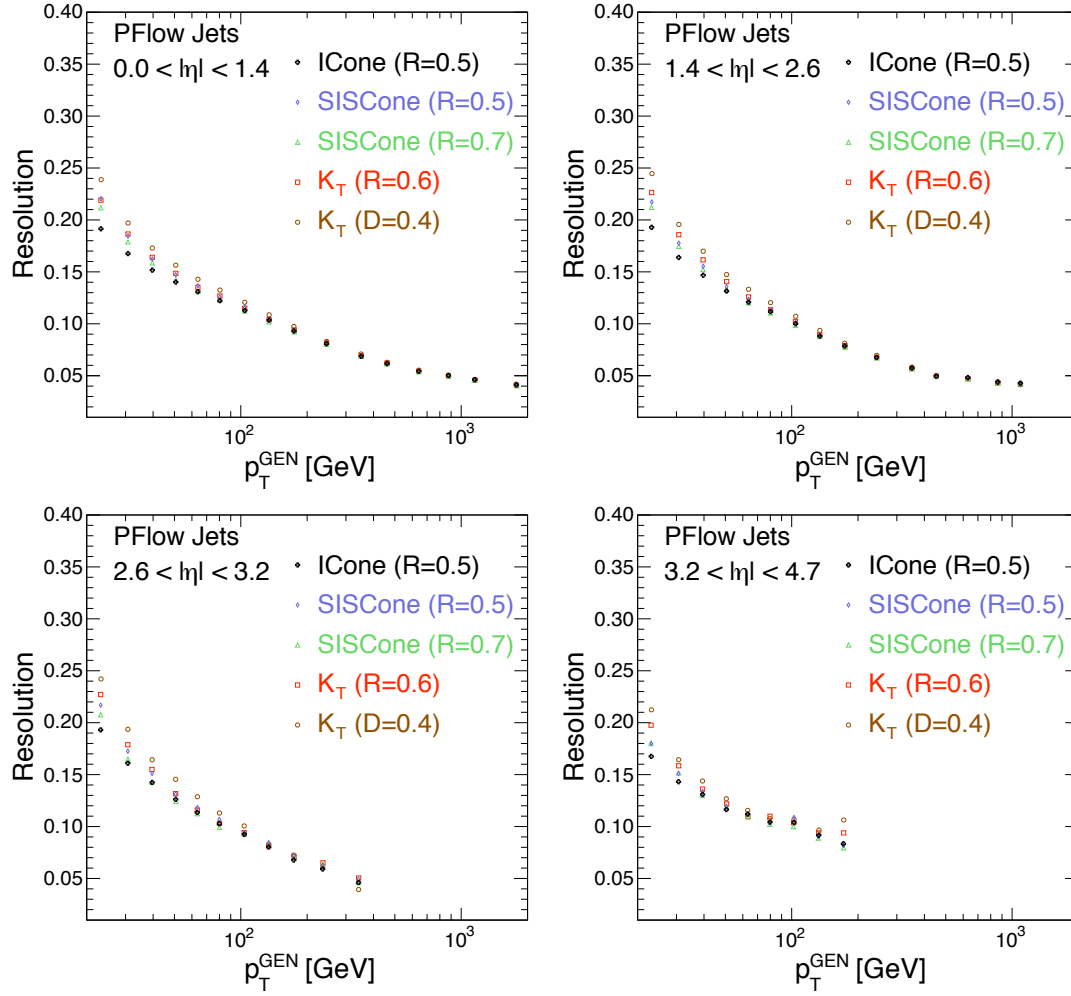


Figure 12: Jet p_T resolution for **PF jets** reconstructed with different jet algorithms and sizes in the barrel (*top left*), endcaps (*top right*), the transition region between endcaps and forward (*bottom left*), and forward (*bottom right*).

274 resolution for all three types of jets, here for Iterative Cone ($R = 0.5$). In the transition region however, the PF jet
 275 η - and ϕ resolution appears to be worse than for calorimeter jets.

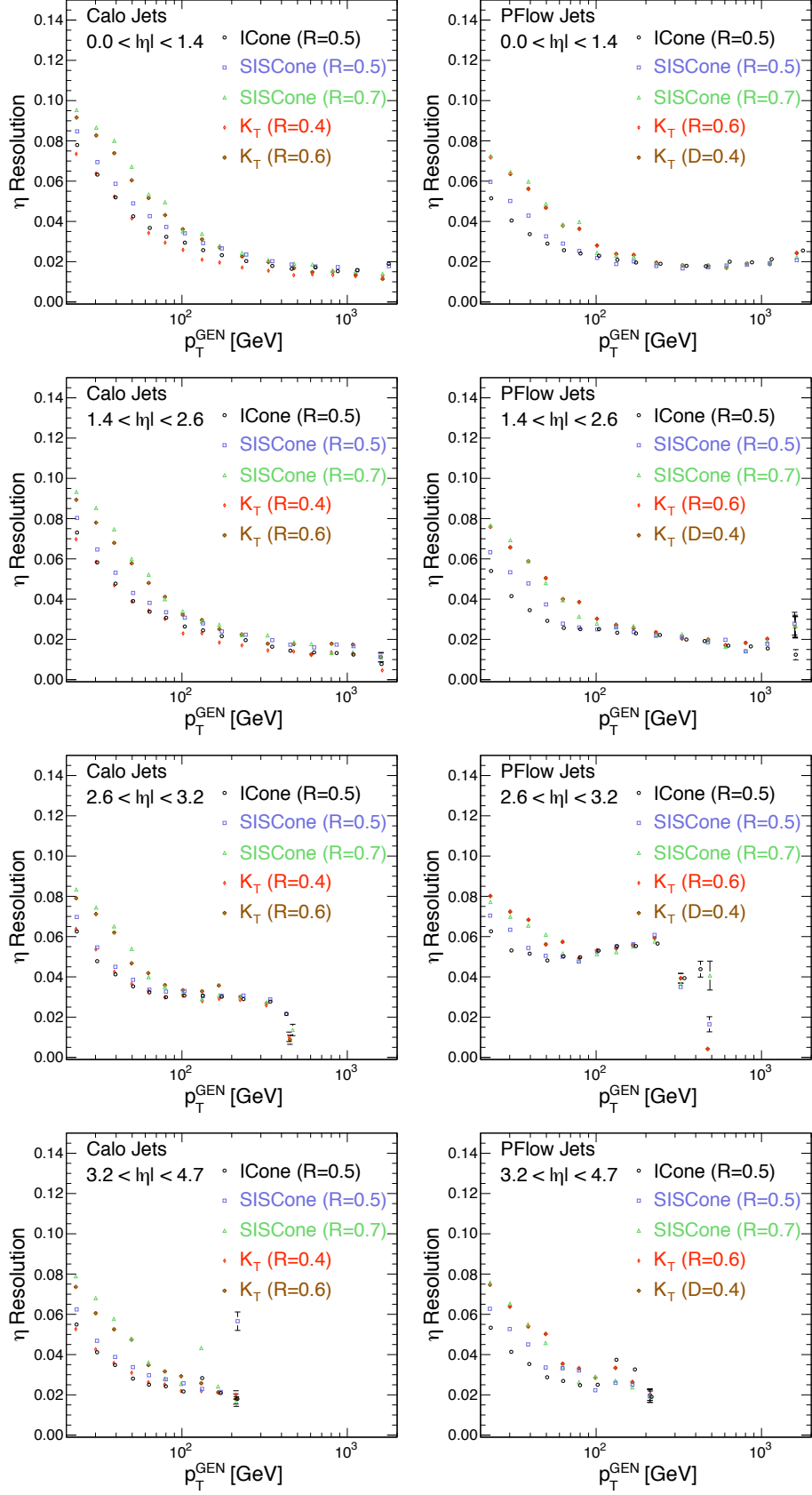


Figure 13: Jet η resolution for calorimeter jets (*left*) and PF jets (*right*) in different detector regions as a function of p_T . From *top* to *bottom*: in the barrel ($0 < |\eta| \leq 1.4$), endcaps ($1.4 < |\eta| \leq 2.6$), transition region between endcaps and forward ($2.6 < |\eta| \leq 3.2$), and forward ($3.2 < |\eta| \leq 4.7$).

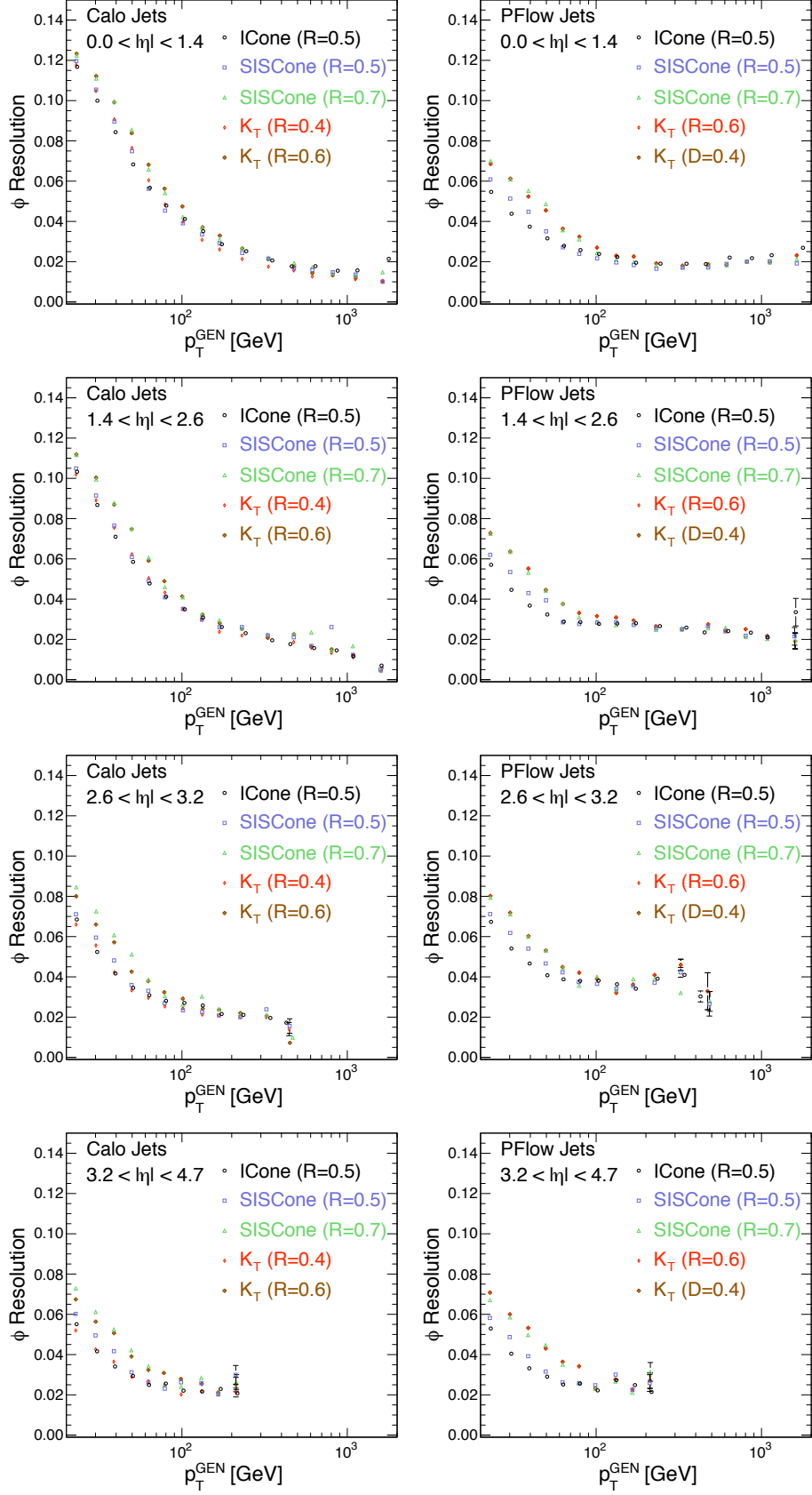


Figure 14: Jet ϕ resolution for calorimeter jets (*left*) and PF jets (*right*) in different detector regions as a function of p_T . From *top* to *bottom*: in the barrel ($0 < |\eta| \leq 1.4$), endcaps ($1.4 < |\eta| \leq 2.6$), transition region between endcaps and forward ($2.6 < |\eta| \leq 3.2$), and forward ($3.2 < |\eta| \leq 4.7$).

Algorithm	η	$a \oplus b \oplus c$
ICone (R=0.5)	$0.0 < \eta < 1.4$	$3.41 \pm 0.042 \oplus 1.30 \pm 0.002 \oplus 0.03 \pm 0.0001$
	$1.4 < \eta < 2.6$	$4.22 \pm 0.045 \oplus 0.96 \pm 0.006 \oplus 0.04 \pm 0.0005$
	$2.6 < \eta < 3.2$	$4.36 \pm 0.073 \oplus 0.39 \pm 0.033 \oplus 0.06 \pm 0.0016$
	$3.2 < \eta < 4.7$	$3.74 \pm 0.025 \oplus 0.00 \pm 0.070 \oplus 0.09 \pm 0.0008$
K_T (R=0.4)	$0.0 < \eta < 1.4$	$5.16 \pm 0.031 \oplus 1.30 \pm 0.003 \oplus 0.03 \pm 0.0001$
	$1.4 < \eta < 2.6$	$5.92 \pm 0.041 \oplus 0.94 \pm 0.007 \oplus 0.03 \pm 0.0007$
	$2.6 < \eta < 3.2$	$4.97 \pm 0.084 \oplus 0.47 \pm 0.033 \oplus 0.06 \pm 0.0019$
	$3.2 < \eta < 4.7$	$4.22 \pm 0.029 \oplus 0.00 \pm 0.068 \oplus 0.09 \pm 0.0010$
K_T (R=0.6)	$0.0 < \eta < 1.4$	$4.03 \pm 0.036 \oplus 1.32 \pm 0.003 \oplus 0.03 \pm 0.0001$
	$1.4 < \eta < 2.6$	$5.38 \pm 0.038 \oplus 0.94 \pm 0.006 \oplus 0.03 \pm 0.0006$
	$2.6 < \eta < 3.2$	$5.47 \pm 0.065 \oplus 0.18 \pm 0.077 \oplus 0.06 \pm 0.0016$
	$3.2 < \eta < 4.7$	$4.54 \pm 0.024 \oplus 0.00 \pm 0.052 \oplus 0.08 \pm 0.0009$
SISCone (R=0.5)	$0.0 < \eta < 1.4$	$4.35 \pm 0.038 \oplus 1.34 \pm 0.003 \oplus 0.03 \pm 0.0001$
	$1.4 < \eta < 2.6$	$5.03 \pm 0.044 \oplus 0.96 \pm 0.007 \oplus 0.04 \pm 0.0006$
	$2.6 < \eta < 3.2$	$4.72 \pm 0.075 \oplus 0.37 \pm 0.037 \oplus 0.07 \pm 0.0016$
	$3.2 < \eta < 4.7$	$4.10 \pm 0.026 \oplus 0.00 \pm 0.065 \oplus 0.09 \pm 0.0009$
SISCone (R=0.7)	$0.0 < \eta < 1.4$	$3.47 \pm 0.045 \oplus 1.37 \pm 0.003 \oplus 0.03 \pm 0.0001$
	$1.4 < \eta < 2.6$	$5.04 \pm 0.042 \oplus 0.98 \pm 0.006 \oplus 0.04 \pm 0.0005$
	$2.6 < \eta < 3.2$	$5.39 \pm 0.066 \oplus 0.10 \pm 0.140 \oplus 0.07 \pm 0.0015$
	$3.2 < \eta < 4.7$	$4.49 \pm 0.023 \oplus 0.00 \pm 0.058 \oplus 0.08 \pm 0.0009$

Table 3: Jet p_T resolution fit parameters for **calorimeter jets**, extracted by fitting the sum of a constant term c , a stochastic term b , and a noise term a to the data points.

Algorithm	η	$a \oplus b \oplus c$
ICone (R=0.5)	$0.0 < \eta < 1.4$	$0.00 \pm 0.031 \oplus 0.99 \pm 0.001 \oplus 0.04 \pm 0.0001$
	$1.4 < \eta < 2.6$	$0.00 \pm 0.108 \oplus 0.92 \pm 0.002 \oplus 0.03 \pm 0.0004$
	$2.6 < \eta < 3.2$	$1.25 \pm 0.225 \oplus 0.88 \pm 0.013 \oplus 0.02 \pm 0.0050$
	$3.2 < \eta < 4.7$	$2.27 \pm 0.172 \oplus 0.52 \pm 0.040 \oplus 0.08 \pm 0.0031$
K_T (R=0.6)	$0.0 < \eta < 1.4$	$0.00 \pm 0.047 \oplus 1.07 \pm 0.001 \oplus 0.03 \pm 0.0001$
	$1.4 < \eta < 2.6$	$2.41 \pm 0.056 \oplus 0.94 \pm 0.005 \oplus 0.03 \pm 0.0006$
	$2.6 < \eta < 3.2$	$3.70 \pm 0.092 \oplus 0.73 \pm 0.018 \oplus 0.04 \pm 0.0022$
	$3.2 < \eta < 4.7$	$4.00 \pm 0.025 \oplus 0.00 \pm 0.114 \oplus 0.09 \pm 0.0008$
K_T (D=0.4)	$0.0 < \eta < 1.4$	$0.00 \pm 0.063 \oplus 1.14 \pm 0.001 \oplus 0.03 \pm 0.0001$
	$1.4 < \eta < 2.6$	$2.80 \pm 0.060 \oplus 0.98 \pm 0.005 \oplus 0.03 \pm 0.0007$
	$2.6 < \eta < 3.2$	$3.33 \pm 0.080 \oplus 0.92 \pm 0.008 \oplus 0.00 \pm 0.1124$
	$3.2 < \eta < 4.7$	$4.31 \pm 0.031 \oplus 0.00 \pm 0.094 \oplus 0.09 \pm 0.0010$
SISCone (R=0.5)	$0.0 < \eta < 1.4$	$0.00 \pm 0.048 \oplus 1.07 \pm 0.001 \oplus 0.03 \pm 0.0001$
	$1.4 < \eta < 2.6$	$1.85 \pm 0.074 \oplus 0.94 \pm 0.005 \oplus 0.03 \pm 0.0006$
	$2.6 < \eta < 3.2$	$2.83 \pm 0.124 \oplus 0.82 \pm 0.017 \oplus 0.03 \pm 0.0028$
	$3.2 < \eta < 4.7$	$2.88 \pm 0.158 \oplus 0.48 \pm 0.049 \oplus 0.08 \pm 0.0034$
SISCone (R=0.7)	$0.0 < \eta < 1.4$	$0.00 \pm 0.045 \oplus 1.03 \pm 0.001 \oplus 0.03 \pm 0.0001$
	$1.4 < \eta < 2.6$	$1.92 \pm 0.064 \oplus 0.91 \pm 0.004 \oplus 0.03 \pm 0.0005$
	$2.6 < \eta < 3.2$	$3.05 \pm 0.100 \oplus 0.72 \pm 0.017 \oplus 0.04 \pm 0.0020$
	$3.2 < \eta < 4.7$	$2.92 \pm 0.131 \oplus 0.49 \pm 0.042 \oplus 0.08 \pm 0.0031$

Table 4: Jet p_T resolution fit parameters for **PF jets**, extracted by fitting the sum of a constant term c , a stochastic term b , and a noise term a to the data points.

Algorithm	η	$a \oplus b \oplus c$
ICone (R=0.5)	$0.0 < \eta < 1.400$	$1.31 \pm 0.062 \oplus 1.02 \pm 0.002 \oplus 0.03 \pm 0.0001$
	$1.4 < \eta < 2.600$	$2.68 \pm 0.050 \oplus 0.91 \pm 0.005 \oplus 0.03 \pm 0.0005$
	$2.6 < \eta < 3.200$	$2.92 \pm 0.104 \oplus 0.70 \pm 0.018 \oplus 0.05 \pm 0.0020$
	$3.2 < \eta < 4.700$	$3.34 \pm 0.027 \oplus 0.00 \pm 0.221 \oplus 0.10 \pm 0.0008$

Table 5: Jet p_T resolution fit parameters for **JPT jets**, extracted by fitting the sum of a constant term c , a stochastic term b , and a noise term a to the data points.

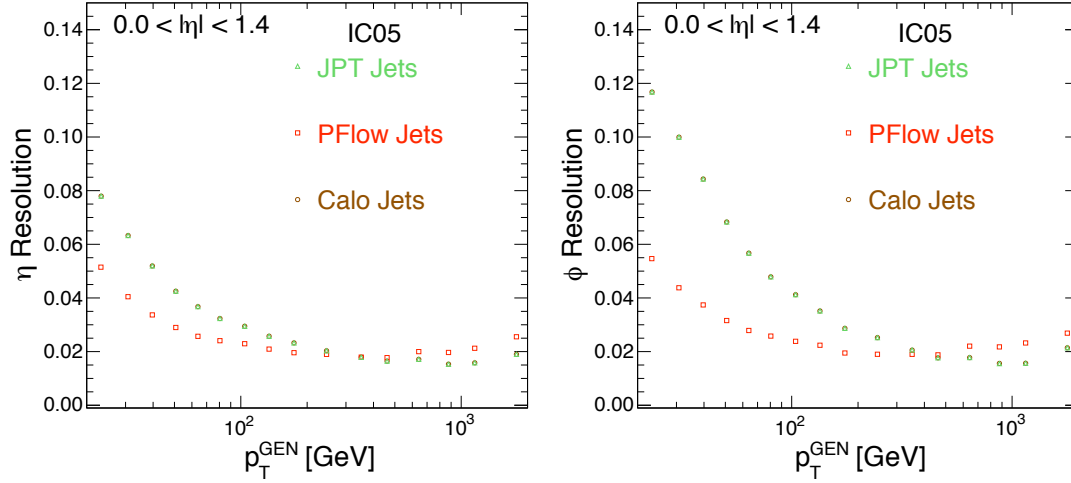


Figure 15: Comparison of η - (left) and ϕ - (right) resolution for calorimeter jets and PF jets, reconstructed with the Iterative Cone ($R = 0.5$) algorithm in the barrel region. JPT jets, which feature identical position resolution as calorimeter jets as explained in the text are also shown.

5.3 The Anti- k_T and Cambridge/Aachen jet algorithms

This section is very preliminary! both the Anti- k_T and Cambridge/Aachen algorithms aren't even available in CMSSW_2.2.X, Table 6 lists the additional software packages and corresponding version tags one needs to compile to enable them. Therefore, we do not plan to have any material concerning these studies publicly approved at this very moment.

Recent publications [8] and very fruitful discussions and exchanges with Gavin Salam however indicate that especially the Anti- k_T algorithm has great potential to behave and perform very similarly to the Iterative Cone algorithm, being infrared- and collinear-safe at the same time! We therefore commit to these studies now, hoping to arrive at a final conclusion and corresponding recommendation when applying these studies to the CMSSW_3.1.X series of our software, which is intended to accompany us into first LHC collisions.

Figures 16, 17, and 18 demonstrate that we can't confirm that especially Anti- k_T jets perform as well in terms of jet p_T resolution at low p_T^{gen} as jets reconstructed with Iterative Cone. There are several caveats however, the most prominent being the jet energy corrections applied to the new algorithms: as they aren't yet officially supported by the CMS Jet Energy Corrections (JEC) group, these corrections needed to be derived from scratch. While trying to follow the exact same methodology used to derive the official corrections for other algorithms, the response as a function of p_T^{gen} in Figure 17 indicates slightly different results. In fact, a quick cross-check reveals better closure for the Anti- k_T response when applying the official Iterative Cone corrections instead of the new dedicated ones. This needs to be further studied.

Furthermore, Gavin Salam has suggested that the differences might arise from the matching requirement: he reports about a quick study using a toy calorimeter simulation where he sees the resolutions converge when tightening the requirement from $\Delta R \leq 0.4$ to $\Delta R \leq 0.2$. We will further investigate the sensitivity of our results to the choice of this cut.

CVS tag	CMSSW package
V01-11-02	RecoJets/Configuration
V02-10-02	RecoJets/JetProducers
V02-09-05	RecoJets/JetAlgorithms
V03-24-03	DataFormats/JetReco
V02-04-00	DataFormats/CaloTowers
V00-11-00	RecoLocalCalo/CaloTowersCreator

Table 6: Additional CMSSW packages and corresponding CVS tags to enable Anti- k_T and Cambridge/Aachen jet reconstruction in CMSSW_2.2.X.

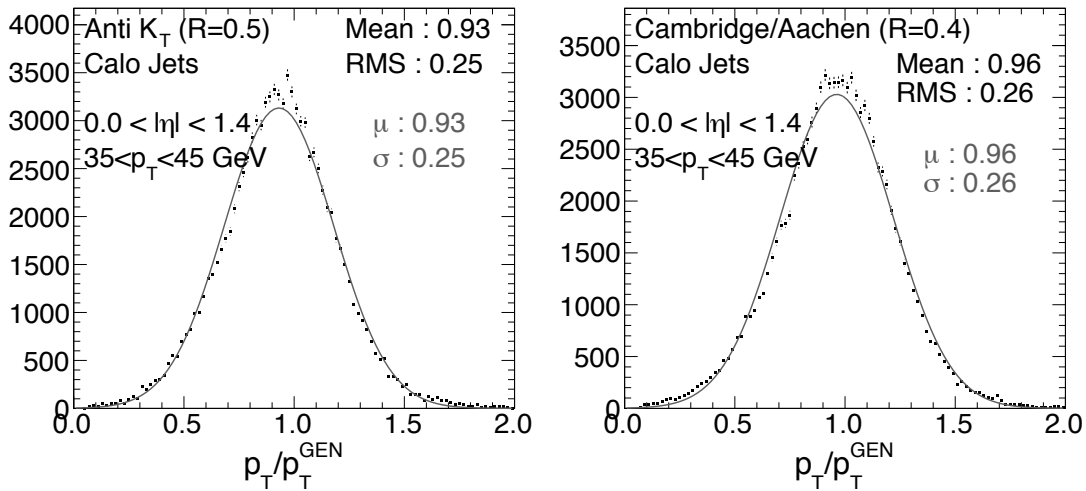


Figure 16: Jet p_T response for calorimeter jets in the range $35 < p_T^{\text{gen}} < 45$ GeV. *Left:* Anti- k_T compared to Iterative Cone ($R = 0.5$). *Right:* Cambridge/Aachen compared to k_T ($D = 0.4$).

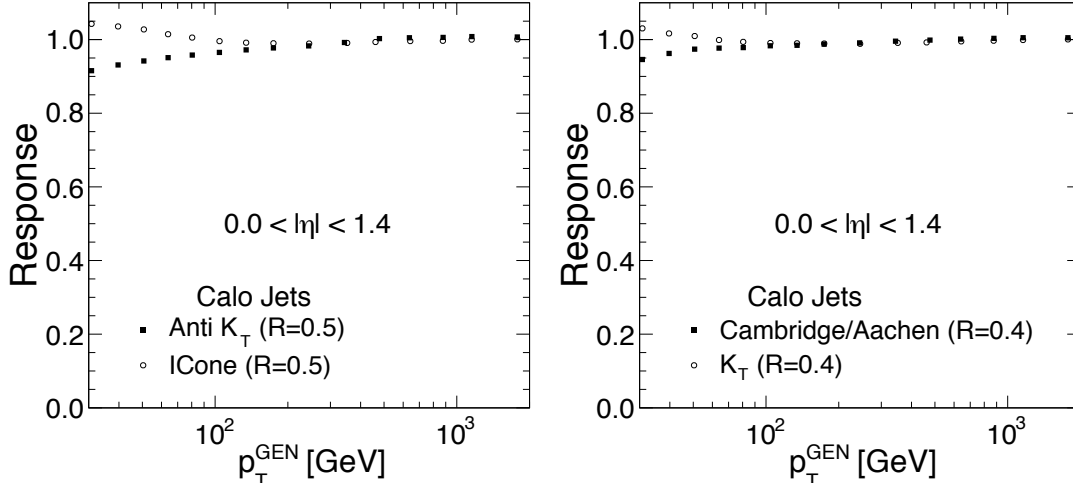


Figure 17: Jet p_T response for calorimeter jets. *Left:* Anti- k_T compared to Iterative Cone ($R = 0.5$). *Right:* Cambridge/Aachen compared to k_T ($D = 0.4$).

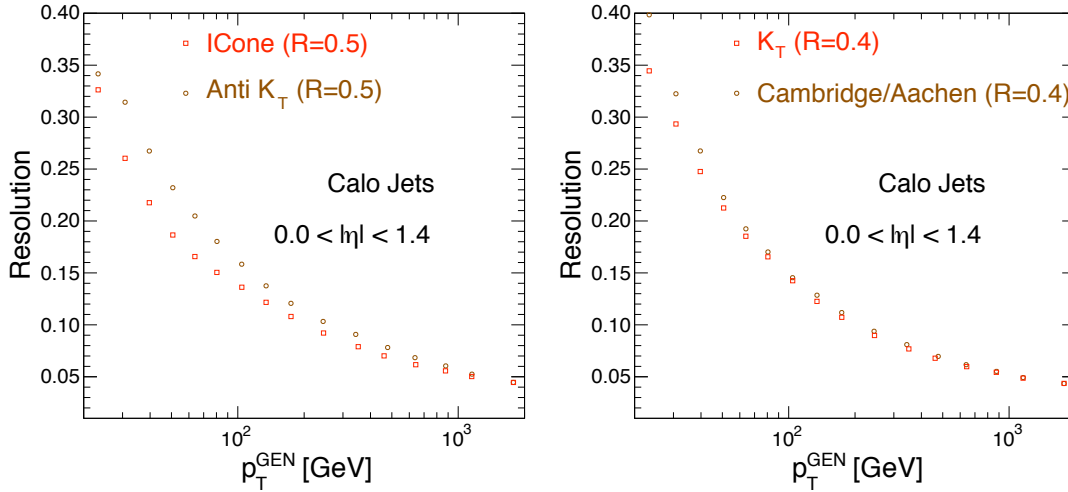


Figure 18: Jet p_T resolution for calorimeter jets. *Left:* Anti- k_T compared to Iterative Cone ($R = 0.5$). *Right:* Cambridge/Aachen compared to k_T ($D = 0.4$).

6 Conclusion

The jet reconstruction performance for all three types of jets available at CMS was presented based on the CMSSW_2_2_X software: calorimeter jets, Particle-Flow (PF) jets, and Jet-Plus-Track (JPT) jets. The response, energy- and position- (η - and ϕ -) resolutions were evaluated and compared for different clustering algorithms, jet size parameters, and detector regions using QCD dijet events. Both PF and JPT jets show significantly improved p_T resolution w.r.t. calorimeter jets by exploiting the excellent performance of the CMS tracking detectors. For PF jets, the tracking information is factored into the reconstruction such that it yields better position resolution as well, something to be expected for JPT jets in future versions of the algorithm. The composition of PF jets was studied in some detail, and evidence was presented that the performance of PF jets depends significantly on the fractional contributions from photons, charged hadrons, and neutrals, which can be exploited in the future e.g. by dedicated jet energy corrections.

We also presented preliminary studies on two jet algorithms, Anti- k_T and Cambridge/Aachen, which will be available to CMS analyzers from the next generation of the CMSSW software (CMSSW_3_1_X). The Anti- k_T algorithm shows similar behavior and performance as the Iterative Cone algorithm while being infrared- and collinear-safe, and we expect to recommend adopting Anti- k_T in favor of Iterative Cone in the near future after further studies.

References

- [1] CMS Collaboration, “*Particle-Flow Event Reconstruction in CMS and Performance for Jets, E_T^{miss} , and $Taus$* ”, CMS PAS **PFT-09-001** (2009).
- [2] T. Sjöstrand, S. Mrenna, and P. Skands, “*PYTHIA 6.4 physics and manual*”, JHEP05:026 (2006).
- [3] R. Demina *et al.*, “*Calorimeter Energy Thresholds for Jet Reconstruction in CMS*”, CMS NOTE 2006-020 (2006).
- [4] V. Chetluru and M. Zielinski, *work in progress*.
- [5] CMS Collaboration, “*Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS*”, CMS PAS **JME-09-002** (2009).
- [6] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, “*Longitudinally invariant K_t clustering algorithms for hadron hadron collisions*”, Nucl.Phys.B**406**:187-224 (1993).
S. D. Ellis and D. E. Soper, “*Successive Combination Jet Algorithm For Hadron Collisions*”, Phys. Rev. D **48** 3160 (1993).
- [7] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, “*Better jet clustering algorithms*”, JHEP 9708:001 (1997).
M. Wobisch and T. Wengler, “*Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering*”, arXiv:hep-ph/9907280v1 (1999).
- [8] M. Cacciari, G. P. Salam and G. Soyez, “*The anti- k_t jet clustering algorithm*”, JHEP0804:063 (2008).
- [9] M. Cacciari and G. P. Salam, “*Dispelling the N^3 myth for the K_t jet-finder*”, Phys.Lett. B**641** 57-61 (2006).
- [10] On the web under: <http://www.lpthe.jussieu.fr/~salam/fastjet/>
- [11] CMS Collaboration, “*CMS physics: Technical design report.*”, CERN-LHCC-2006-001 (2006).
- [12] G. P. Salam and G. Soyez, “*A practical seedless infrared-safe cone jet algorithm*”, JHEP05(2007)086 (2007).
- [13] CMS Collaboration, “*Plans for Jet Energy Corrections at CMS*”, CMS PAS **JME-07-002** (2007).
- [14] CMS Collaboration, “*Offset Energy Correction for Jets*”, CMS PAS **JME-09-003** (2008).
- [15] CMS Collaboration, “*Determination of the Relative Jet Energy Scale at CMS from Dijet Balance*”, CMS PAS **JME-08-003** (2008).
- [16] CMS Collaboration, “*Jet energy calibration with photon+jet events*”, CMS PAS **JME-09-004** (2009).
CMS Collaboration, “*Jet energy correction using $Z \rightarrow ee + \text{Jet } p_T$ balance*”, CMS PAS **JME-09-005** (2009).
CMS Collaboration, “*Jet energy correction using $Z \rightarrow \mu\mu + \text{Jet } p_T$ balance*”, CMS PAS **JME-09-009** (2009).

- 344 [17] F. Pandolfi and D. del Re, *work in progress*.
- 345 [18] CMS Collaboration, “*The 2008 CMS Computing, Software and Analysis Challenge*”, CMS IN **2008/044**
346 (2008).
347 Further documentation on the web under:
348 <https://twiki.cern.ch/twiki/bin/view/CMS/CSA08AnaJets>