Chapter 1

Jets at LHC

Contributors: Daniele Benedetti (INFN Perugia), Andrea Giammanco (UCL Louvain), Chiara Roda (INFN Pisa), Attilio Santocchia (INFN Perugia), Iacopo Vivarelli (INFN Pisa)

1.1 Introduction

**** PER MATTEO: LA PARTE DA RISCRIVERE INIZIA QUI

In this chapter different aspects of jets will be discussed and analysed. In particular here we concentrate on the algorithmic task of clustering the input objects (e.g. simulated particles, calorimeter cells, tracks) into jets, on the jet calibration, on the task of assigning a jet to a specific parton, and on the "Energy Flow" (or "Particle Flow") method.

1.2 Jet clustering

1.2.1 Algorithms

The following jet reconstruction algorithms will be mentioned in this chapter, although many others exist: the *iterative cone* algorithm (IC), the inclusive k_T algorithm (k_T) and the *MidPoint Cone* algorithm (MC) [5].

In the **iterative cone algorithm**, an E_T-ordered list of input objects (particles or calorimeter towers) is created. A cone of size R in η, ϕ space is cast around the input object having the largest transverse energy above a specified seed threshold. The objects inside the cone are used to calculate a proto-jet direction and energy. The computed direction is used to seed a new proto-jet. The procedure is repeated until stability is reached (i.e. the energy of the proto-jet changes by less than 1% between two consecutive iterations and the direction of the proto-jet changes by $\Delta R < 0.01$). When a stable proto-jet is found, all objects in the proto-jet are removed from the list of input objects and the stable proto-jet is added to the list of jets. The whole procedure is repeated until the list contains no more objects with an E_T above the seed threshold. The cone size and the seed threshold are tunable parameters of the algorithm.

The **inclusive k_T jet algorithm** is a cluster-based jet algorithm. The cluster procedure starts with a list of input objects, stable particles or calorimeter cells. For each object i and each pair (i, j) the following distances are calculated:

$$
\begin{aligned}
d_i &= (E_{T,i})^2 R^2 \\
d_{i,j} &= min(E_{T,i}^2, E_{T,j}^2)\Delta R_{i.j}^2 \quad \text{with} \quad \Delta R_{i.j}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2
\end{aligned}
$$

where R^2 is a dimensionless arbitrary parameter. The algorithm searches for the smallest d_i or d_{ij}. If a value of type d_{ij} is the smallest, the corresponding objects i and j are removed from the list of input objects. They are merged using one of the recombination schemes and filled as one new object into the list of input objects. If a distance of type d_i is the smallest, then the corresponding object i is removed from the list of input objects and filled into the list of final jets. The procedure is repeated until all objects are included in jets. The algorithm successively merges objects which have a distance $R_{ij} < R$. It follows that $R_{ij} > R$ for all final jets i and j.

The **midpoint-cone algorithm** was designed to enforce the splitting and merging of jets. It also uses an iterative procedure to find stable cones (proto-jets) starting from the cones around objects with an E_T above a seed threshold but, contrary to the iterative cone algorithm described above, no object is removed from the input list. This can result in overlapping proto-jets (a single input object may belong to several proto-jets). Then, in order to ensure the collinear and infrared safety of the algorithm, a second iteration of the list of stable jets is done. For every pair of proto-jets with distance less than the cone diameter, a midpoint is calculated as the direction of the combined momentum. All these midpoints are then used as additional seeds to find more proto-jets. When all proto-jets are found, the splitting and merging procedure is applied, starting with the highest E_T proto-jet. If a proto-jet does not share objects with other proto-jets, it is defined as a jet and removed from the proto-jet list. Otherwise, the transverse energy shared with the highest E_T neighbouring proto-jet is compared to the total transverse energy of this neighbour proto-jet. If the fraction is greater than an arbitrary threshold f (typically 50%) the proto-jets are merged, otherwise the shared objects are individually assigned to the closest proto-jet. The procedure is iterated, always starting from the highest E_T proto-jet, until no proto-jets are left. The parameters of the algorithm include a seed threshold, a cone radius, the threshold f mentioned above, and also a maximum number of proto-jets that are used to calculate midpoints.

1.2.2 Particle Jets

Here we call "particle jets" those obtained in simulated data by applying the jet clustering algorithms to all stable particles (charged and neutral) as obtained at the generator level after the hadronization step, without considering any of the detector effects (like calorimeter resolution or the sweeping from the magnetic field[1]). A particle jet includes any particle emerging from the hard scattering process or from the underlying event. The jet clustering algorithm is applied to the particle four momenta at the interaction vertex. Some authors exclude the neutrinos from the list of input particles, since they cannot give a signal in the detector, not even in principle.

**** PER MATTEO: LA PARTE DA RISCRIVERE FINISCE QUI (se ricordo bene)

1.2.3 Calorimeter Jets

The calorimeter jets, or reconstructed jets[2], are obtained by applying the jet clustering algorithm to the calorimeter signals. Calorimeter signals are defined by grouping the calorimeter cells to obtain a granularity best suited to the scale of hadronic showers. Noise reduction is also applied at this stage.

The most common clusterization consists in assembling calorimeters cells into towers in (η, ϕ) space. CMS builds towers of dimension $(\Delta\eta \times \Delta\phi) = 0.087 \times 0.087$ (the granularity of the hadronic section) in the central region, gradually increasing in the end-cap and forward region, for a total of 4167 towers. The noise suppression algorithm consists in building the towers using only those cells whose signals is higher than a predefine energy threshold, whose value depends on η and on the longitudinal sample. Various threshold schemes have been considered, and the most used so far in the analyses uses 0.7 GeV and 0.85 GeV thresholds for the HB and HO respectively. In this scheme the noise contribution for a $\Delta R = 0.5$ cone jet is equal to 1.4 GeV with a negligible loss of signal.

[1]The minimum transverse momenta required to reach the calorimeter inner surface is about 350 MeV for the ATLAS system and about 700 MeV for the CMS system.

[2]Although it has to be reminded that jets can be formed from other inputs, e.g., the "Energy Flow objects", see Sec. 1.5.

In ATLAS 6400 towers are built with a fixed dimension of $(\Delta\eta \times \Delta\phi) = 0.1 \times 0.1$, corresponding to the granularity of the central hadronic section. The high number of cells in the electromagnetic sections and the level of noise does not allow to apply an asymmetric noise cut on cells since it would introduce a positive bias in the measured signal. Towers are thus built without any noise suppression. A second clusterization scheme is developed to obtain better noise suppression while avoiding any large bias. This scheme consists of building three-dimensional clusters associating neighboring cells which belong to any calorimeter section[1], with three minimum cell thresholds: one to start the search (T_{seed}), one to expand the cluster (T_{neigh}) and a third one to append contour cells (T_{cont}). The defaults threshold values, applied to the absolute cell energy, are $T_{seed} = 4\sigma_{noise}$, $T_{neigh} = 2\sigma_{noise}$, $T_{cont} = 0\sigma_{noise}$. At last, a splitting procedure is applied to separate superimposed or connected clusters, based on the presence of multiple local signal maxima in one cluster. A large reduction of noise is obtained if three-dimensional clusters are used instead of the towers.

1.3 Optimization of the clustering algorithms

This section summarizes the studies of Ref. [2] on the jet clustering optimization. This optimization is defined in terms of quality criteria or quality markers, related to the reconstruction efficiency of the complete kinematics of the primary quark event topology.

Physics effects like pile-up, underlying event and radiation enlarge this mean error. The scope of this study is to find the most efficient jet finding setup in the presence of these effects, in order to maximise the fraction of events for which all quarks are matched to reconstructed jets, according to some predefined criteria. Hence, events suffering from a large amount of hard gluon radiation will be rejected.

This study has been performed with simulated particle information as input to the jet finding algorithms, and deals with algorithmic and physics effects, independently of detector specificities. It has to be kept in mind that instrumental effects can, in principle, alter significantly the conclusions of this study. Work is currently in progress in CMS for an analogous study with the full detector simulation and reconstruction chain.

For all algorithms generated and stable final state particles are used as input. Muons and neutrinos are excluded, and the effects of the magnetic field are not taken into account. All particles are assumed to emerge from the primary vertex, where the clustering is performed.

1.3.1 Event generation

For this study, processes with two, four, six and eight primary quarks in the final state (dileptonic and single-leptonic top decays in $t\bar{t}$ events, single-leptonic and fully hadronic top decays in $t\bar{t}H$) have been considered.

Proton collisions at 14 TeV have been generated at a luminosity of 2×10^{33} cm^{-2}s^{-1}. The $t\bar{t}$ events were generated using PYTHIA version 6.2 [3] and the $t\bar{t}H$ events were generated with compHEP version 41.10 [4], interfaced to PYTHIA version 6.215 for showering and hadronization. For the leptonic decays, only electrons and muons are considered.

1.3.2 Event selection and jet-quark matching

A realistic event selection (inspired by $t\bar{t}$ and $t\bar{t}H$ analyses) is applied. The reconstructed jets are required to have a transverse energy larger than 20 GeV, and to be within the tracker acceptance required for a proper b-tagging performance ($|\eta| < 2.4$) in the CMS experiment. Isolated signal leptons from the W-decay are removed from the jet finding input. Only if the number of jets passing these criteria is larger than or equal to the number of primary partons the event is considered for the analysis.

An iterative procedure is used to match the reconstructed jets to the generated quarks based on the ΔR distance in the (η,ϕ) plane. For each possible jet-quark couple the ΔR-value is calculated, and the smallest value is considered as a correct jet-quark matching and is removed from the list for the next iteration. When more jets have a minimal ΔR-value with the same quark, the couple with the lowest ΔR-value is taken. This procedure is iterated until all jets have their respective quark match.

1.3.3 Description of the quality markers

In order to obtain an efficient reconstruction of the kinematics of the primary partons, the selected jets should match both in energy and direction the primary partons. Variables called quality markers are defined to quantify the goodness of the event reconstruction from that perspective. Although physics effects of pile-up, gluon radiation and underlying event will degrade the overall event reconstruction efficiency, it has to be reminded that in principle they can affect differently the considered jet definitions.

Event selection efficiency "ϵ_s"

This efficiency is defined as the fraction of events that pass the event selection i.e. the events with a number of jets greater than the number of partons with

$E_T > 20$ GeV and $|\eta| < 2.5$. When the selection is applied on quark level, the efficiency is equal to 80% for the two quarks final state, 62% for the four quarks final state, 61% for the six quarks final state and 52% for the eight quarks final state.

Angular distance between jet and parton "Frac α_{jp}^{max}"

A jet is considered to be well reconstructed, if the ΔR distance between its direction and its best matched quark direction, α_{jp}, is sufficiently small. For each event, this results in a list of increasing α_{jp}^i-values, $\{\alpha_{jp}^1, ..., \alpha_{jp}^n = \alpha_{jp}^{max}\}$, where n is the amount of primary quarks in the considered event topology. Hence, α_{jp}^{max} is defined as the maximum α_{jp}^i-value of all i jet-quark pairs in the event. The α_{jp}^i distributions for a four quarks final state are shown in Fig. 1.1.

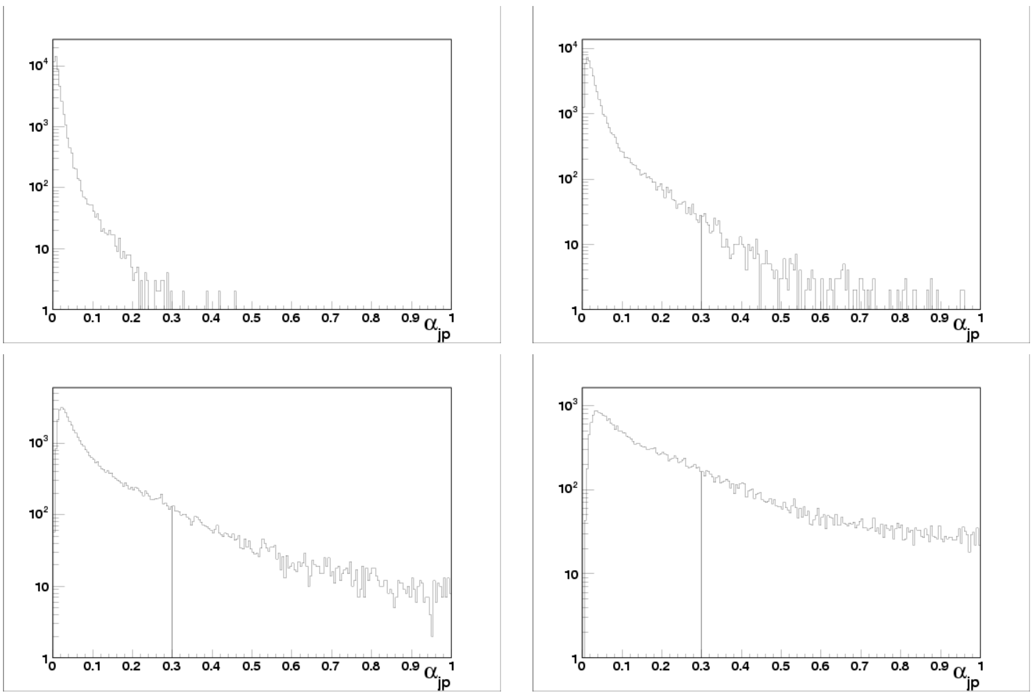

Figure 1.1: Distributions of α_{jp}^i in increasing order for the IC algorithm with a cone radius of 0.4 in the case of a final state with four quarks. The 0.3 rad cut as discussed in the text is indicated.

The last of these plots represents the α_{jp}^{max} variable. To quantify the angular reconstruction performance of a particular jet definition, a quality marker is defined as the fraction of events with a α_{jp}^{max} value lower than 0.3 and denoted as "Frac α_{jp}^{max}".

Energy difference “Frac β_{jp}^{max}”

The reconstructed energy of the primary quarks is usually biased and has a broad resolution. Figure 1.2 (left) shows the average fraction of the quark energy that is reconstructed for a specific algorithm configuration as a function of the reconstructed transverse jet energy. Such a calibration curve can be interpreted as an estimator for the expected reconstructed energy. For this plot only well matched (α_{jp} <0.3), non-overlapping jets were taken into account. For the iterative cone algorithm, a jet is considered to be non-overlapping, if its ΔR distance to any other jet is larger than twice the value of the cone radius parameter of the algorithm. It is the aim of jet calibration studies to determine these average corrections to be applied on the reconstructed jet energies. Therefore the remaining component is the energy resolution.

The β_{jp}^{i} values are defined for each primary quark i as the distance from the expected energy fraction (deduced from the fitted function in Fig. 1.2 left) in units of standard deviations. For each selected event, the primary quark with the highest β_{jp}^{i} value, called β_{jp}^{max} is considered to be the one with the worst reconstruction performance from the energy point of view. An example for the β_{jp}^{max} distribution is shown in Fig. 1.2 (on the right). An energy related quality marker is defined as the fraction of events with a β_{jp}^{max} lower than 2 standard deviations, and denoted as “Frac β_{jp}^{max}”.

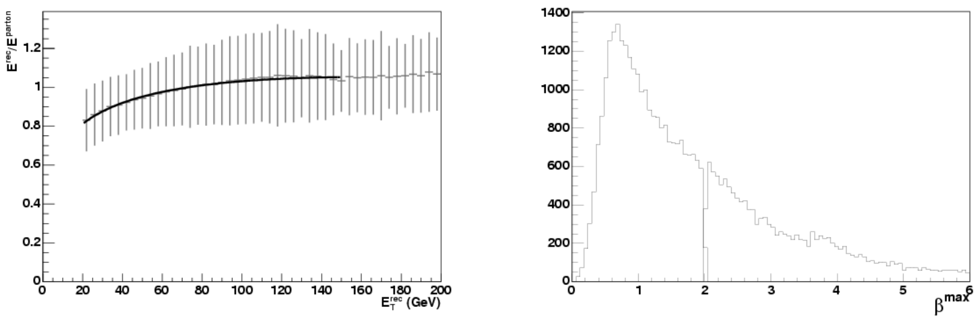

Figure 1.2: Left: example of a $\frac{E^{jet}}{E^{parton}}$ vs. E_T^{jet} curve for the IC algorithm with a cone radius of 0.4, applied on a final state with four primary quarks. The vertical bars illustrate the resolution. Right: distribution of β_{jp}^{max} for the IC algorithm with a cone radius of 0.4, applied on a final state with four primary quarks.

Combined variable "Frac(α_{jp}^{max}+β_{jp}^{max})"

A combined variable is defined as the fraction of events in which both the direction and the energy of the n primary quarks are well reconstructed following the definitions described above. The correlation between α_{jp}^{max} and β_{jp}^{max} is shown in Fig. 1.3 (left), where both quality criteria define a rectangular area in which the kinematics of the primary quarks are sufficiently well reconstructed from the analysis performance point of view. As an illustration of the separation power of this combined variable, the reconstructed spectrum of the hadronic top quark mass in the semileptonic $t\bar{t}$ final state is shown in Fig. 1.3 (right). The black histogram refers to the events in which the jets are reconstructed with $\alpha_{jp}^{max} < 0.3$ and $\beta_{jp}^{max} < 2$ (events inside the box of Fig. 1.3 left). The grey histogram refers to the events in which the kinematics of the primary quarks are badly reconstructed based on the combined variable (events outside the box of Fig. 1.3 left).

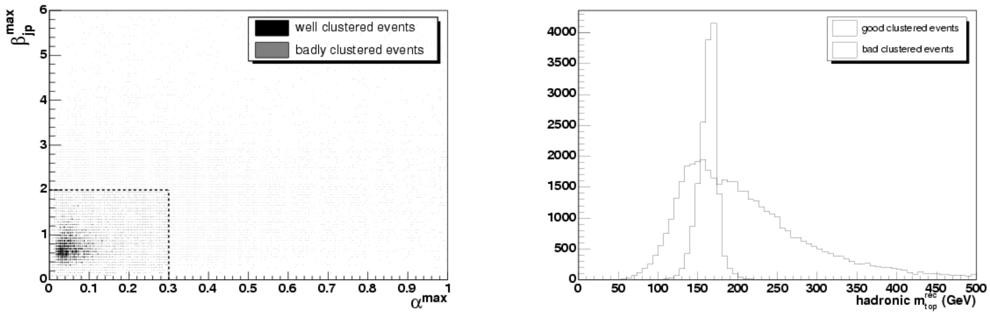

Figure 1.3: Left: box plot of β_{jp}^{max} vs. α_{jp}^{max} for the IC algorithm with a cone radius of 0.4, applied on a final state with four primary quarks. Right: distribution of the hadronic top quark mass, using jets clustered with the IC algorithm with a cone radius of 0.4, applied on a final state with four primary quarks.

Overall quality marker "FracGood"

The fraction of selected and well reconstructed events, i.e. the selection efficiency ϵ_s, multiplied by the combined variable Frac(α_{jp}^{max}+β_{jp}^{max}) is defined as "FracGood".

This last quality marker is interpreted as an estimate for the reconstruction efficiency of the kinematics of the primary quarks of the complete event, and therefore used to compare different algorithms and setups. Although this variable gives a powerful overall indication of a reasonable jet definition,

	IC jet radius		k_T R-parameter		MC jet radius		MC Overlap Threshold	
	Value	*FracGood*	*Value*	*FracGood*	*Value*	*FracGood*	*Value*	*FracGood*
2 quarks	0.5	53.9	0.6	54.9	0.5	42.4	0.40	40.3
4 quarks	0.5	22.3	0.5	23.8	0.3	22.8	0.40-0.50	22.9
6 quarks	0.3	11.2	0.4	12.9	0.2	12.1	0.50-0.60	11.8
8 quarks	0.3	4.85	0.3	5.93	0.2	5.72	0.60	5.0

Table 1.1: Overview of the optimal parameter values with their respective estimate of the fraction of well reconstructed events.

it is sometimes useful to consider the partial information of the individual quality markers. Depending on the priorities of a specific physics analysis, one would be interested in the average number of reconstructed jets, or the energy resolution for non-overlapping jets, or the efficiency of the angular matching between primary quark and jet. The average number of jets gives an idea of the sensitivity to pile-up, underlying event, and the rate of fake jets, while the energy resolution can be linked to the issue of jet calibration.

1.3.4 Results

Table 1.1 summarizes the optimal parameter values for the three jet clustering algorithms, and for each of the considered event topologies. For each optimal jet configuration, the respective estimate of the fraction of well reconstructed events is given.

Robustness of the method against hard radiation

The sensitivity of the overall observations to the radiation of gluons with a large transverse momentum relative to their mother quark, or from the initial state proton system, is investigated in the following. The distributions of the α^i_{jp}-values ordered by their magnitude within an event are shown in Fig. 1.4 for a sample without initial and final state radiation[3].

This has to be compared directly to Fig. 1.1 which shows the same plots including final state radiation. Obviously, the long tails are not present in the case without radiation which indicates that the ΔR cut of 0.3 for the worst jet is not expected to have an effect in this case. The observation is indeed, that the Frac(α^{max}_{jp}+β^{max}_{jp}) quality marker has a flat distribution, but not the selection efficiency and therefore the "FracGood" quality marker.

Fig. 1.5 (left) shows the fraction of selected, well clustered semileptonic $t\bar{t}$ events with and without initial and final state radiation for the *Iterative Cone*

[3]PYTHIA parameters $MSTP$ 61 and 71 were switched off.

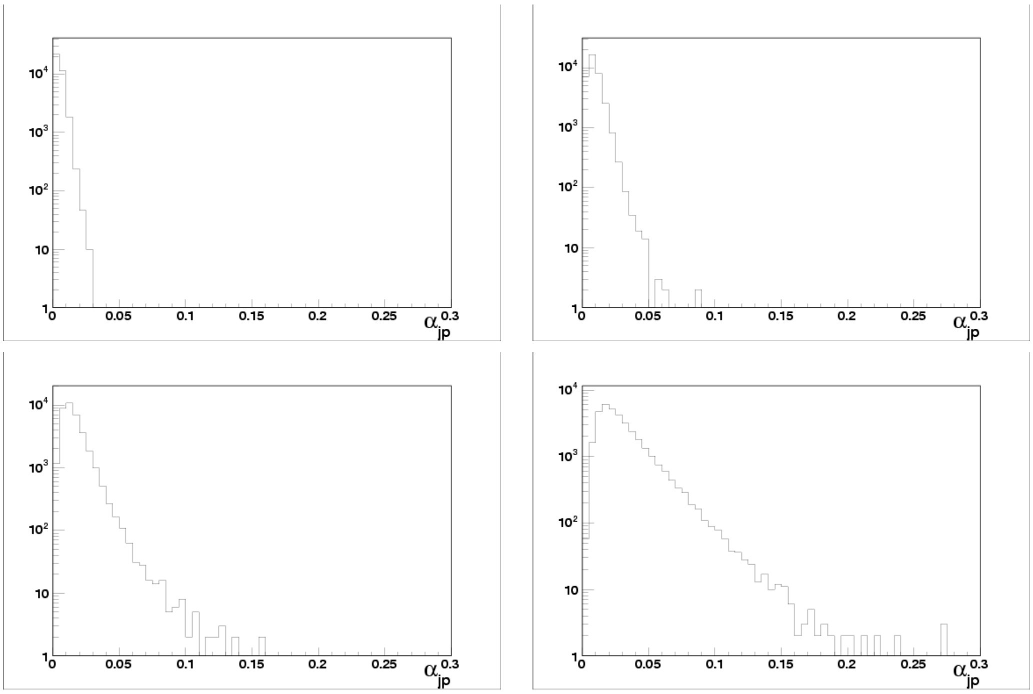

Figure 1.4: Distributions of α^i_{jp} in increasing order of magnitude for the IC algorithm in the case of a final state with four primary quarks which do not radiate hard gluons.

algorithm. The addition of radiation results in an overall lower efficiency, but the optimal cone radius and the shape of the curve are robust. A similar observation was obtained for the inclusive k_T algorithm in Fig. 1.5 (right).

In order to quantify the effect of radiation on the resolutions, Fig. 1.6 shows the two cases for the *Iterative Cone* and the inclusive k_T algorithm for the case with four partons in the final state. The energy resolution is defined as the RMS divided by the mean value of the E^{jet}/E^{quark} distribution, and the angular resolution is defined by the width of a gaussian fit to the symmetrized ΔR distribution. As expected, the overall resolutions are better in the case without radiation, but the shape of the curves remains invariant.

1.4 Jet Calibration

The goal of jet calibration is to correct for various effects that degrade the measurement of the jet energy in the calorimeter. These effects may be divided in two classes : detector driven effects (noise, non-compensation,

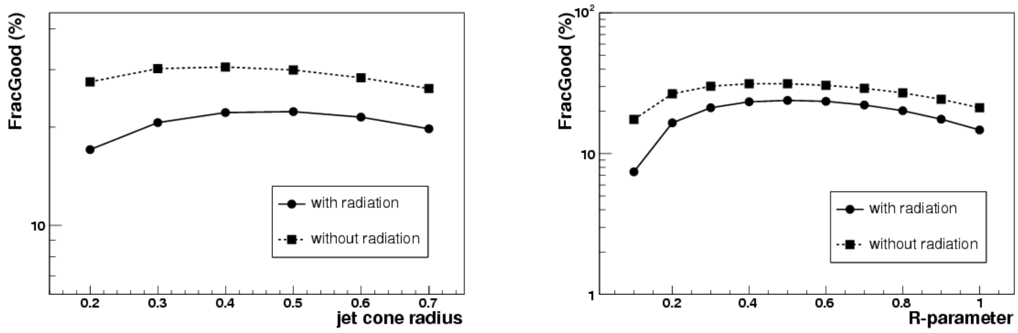

Figure 1.5: Left: influence of hard gluon radiation on the fraction of selected, well clustered events, as a function of the IC cone radius in the case with four primary quarks in the final state. Right: influence of hard gluon radiation on the fraction of selected, well clustered events, as a function of the k_T R-parameter in the case with four primary quarks in the final state.

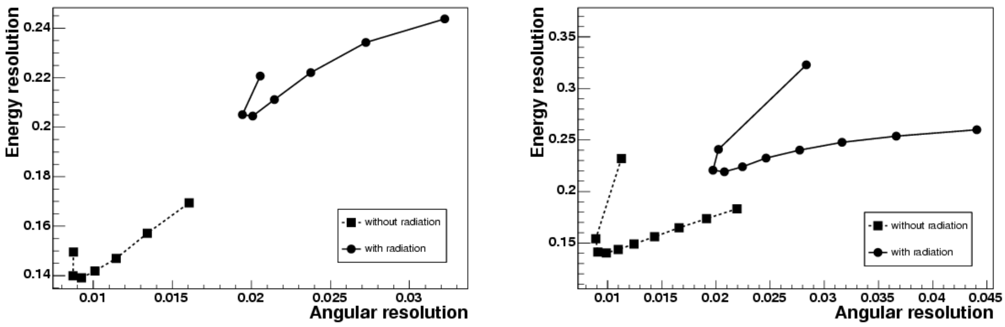

Figure 1.6: Energy resolution versus angular resolution (ΔR distance between jet and quark) for the IC algorithm (left) and k_T algorithm (right) in the case of four jets in the final state.

cracks, dead-material, magnetic field effects, pile-up) and physics driven effects (underlying event, showering effects, clustering). Many different strategies may be chosen to implement the jet calibration and to check its performance and systematics. In the next subsections the baseline strategies for the two experiments are discussed.

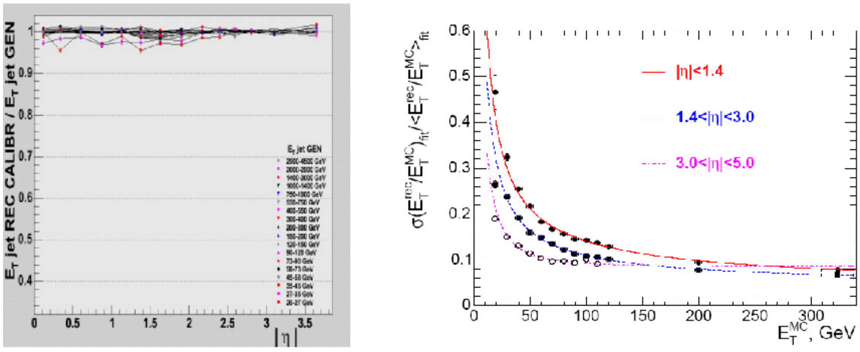

Figure 1.7: CMS Jet linearity after applying calibration (left) as a function of the particle jet pseudo-rapidity and in various particle jet energy ranges. Jet energy resolution resolution (right) as a function of particle jet energy in three ranges of pseudo-rapidity. Jets have been reconstructed with the IC algorithm with $\Delta R = 0.5$ [7].

1.4.1 Calibration to the Particle Jet

The degradation of the jet measurement performance caused by the detector effects may be corrected by applying weights that calibrate the reconstructed jet to the particle jet. The idea to separate detector and physics effect corrections is based on the fact that these two classes of effects have different correlation to the jet kinematics.
In order to obtain the calibration parameters, both ATLAS and CMS use QCD di-jet events generated with PYTHIA [3] and simulated with the full detector descriptions. Calorimeter and particle jets are matched on the base of their distance in the (η, ϕ) space.

In CMS the pseudo-rapidity range $|\eta| < 4.8$ is divided into 16 regions. For each region the mean ratio of reconstructed jet transverse energy (E_T^{calo}) to particle jet transverse energy (E_T^{ptcl}), $R_{jet} = E_T^{calo}/E_T^{ptcl}$, as a function of E_T^{ptcl}, is approximated by a set of functions [15]. The values of R_{jet} obtained are then used to correct the transverse jet energy. Since R_{jet} is a function of E_T^{ptcl}, which is unknown in real data, an iterative procedure is used to obtain for each calorimeter jet energy the best estimate of the calibration

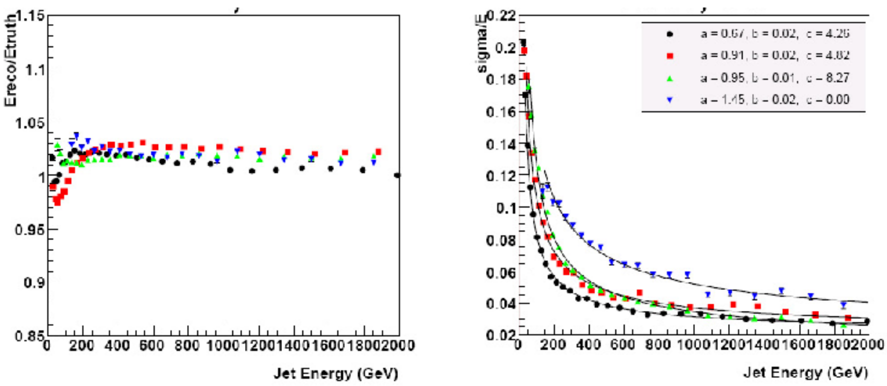

Figure 1.8: ATLAS jet linearity (left) and resolution (right) after applying calibration as a function of the particle jet energy and in various pseudo-rapidity ranges ($|\eta| < 0.7$ (black circles), $0.7 < |\eta| < 1.5$ (red squares), $1.5 < |\eta| < 2.5$ (green triangles), $2.5 < |\eta| < 3.2$ (blue triangles)). Jets have been reconstructed with the $\Delta R = 0.7$ cone algorithm.

parameter [7]. The linearity and the resolution obtained by applying this calibration to a statistical independent sample of QCD di-jet events are shown in Figure 1.7. The maximum deviation from linearity for the E_T range [20 GeV - 4 TeV] is $\sim 5\%$. The energy resolution in the region $|\eta| < 1.4$ is :

$$\frac{\sigma(E_T)}{E_T} = \frac{1.25}{\sqrt{E_T(GeV)}} \oplus \frac{5.6}{E_T(GeV)} \oplus 0.03 \tag{1.1}$$

.
In ATLAS the calibrated jet energy is obtained by applying the weights (w_i) to the cell energies (E_{cell}) that compose the jets:

$$E^{calib} = \sum_i w_i E_i \tag{1.2}$$

The weights, which depend on the position and energy density of the cells, are extracted by minimizing a χ^2 defined as :

$$\chi^2 = \sum_j \left(\frac{E_j^{calib}}{E_j^{ptcl}} - 1 \right)^2 \tag{1.3}$$

where the index j runs on the ensemble of jets of all the events. The dependence of the weight w_i on the cell energy density is parameterized with a polynomial. The basic idea behind this kind of calibration, which exploits the shower shapes, is that hadronic showers are diffuse while electromagnetic ones are dense. Therefore w_i is typically larger than 1 for low cell energy densities and is around 1 for high cell energy densities. In order to partially decouple the effect of cracks from other detector effects the weights w_i are calculated selecting only jets that are far from cracks. The residual non-linearities induced by cracks are than corrected applying an additional factor obtained from signal linearity requirements. The linearity and resolution, as a function of the particle jet energy, obtained on a sample of QCD di-jet events for various pseudo-rapidity regions are shown on figure 1.8. The maximum deviation from linearity is within 2% in the jet energy range [40 GeV - 2 TeV] and the resolution in the pseudo-rapidity region $|\eta| < 0.7$ is equal to :

$$\frac{\sigma(E)}{E} = \frac{0.67}{\sqrt{E(GeV)}} \oplus \frac{4.3}{E(GeV)} \oplus 0.02 \tag{1.4}$$

The jet linearity, as estimated using a sample of events with different parton composition and topology, generated by HERWIG [16], is also well within $\pm 2\%$.

1.4.2 Parton-level calibration

Calibration to the parton jet can be implemented as a second step in addition to particle jet calibration or as a single step which corrects for both detector and physics effect. ATLAS is presently considering the first strategy, while CMS has implemented both [17].

The definition of the parton jet energy is somehow artificial, since partons cannot be defined as isolated objects (not even in the short time scales of the hard interactions). However, results obtained by previous experiments [18] show how this calibration can help understanding the many effects present in hadronic interactions.

A first difference between particle and parton jet is caused by the smearing produced during final state radiation and fragmentation. Both phenomena generate particles which may not be clustered into the particle jet. This results in a fraction of the parton jet energy not attributed to the particle jet. In the case of cone clustering algorithms these losses are indicated as out-of-cone losses. Second, some of the particles generated in the underlying event may fall in the jet region and be attributed to the particle jet although this contribution is not related to the parent parton jet. In this section some possible strategies to correct for these effects are discussed.

A first possibility, exploited by CMS, to obtain the parton jet energy scale is to use simulated events and obtain a calibration constant $k_{ptcl} = p_T^{ptcl}/p_T^{parton}$ as a function of the transverse energy of the parton. In figure 1.9 (left) the

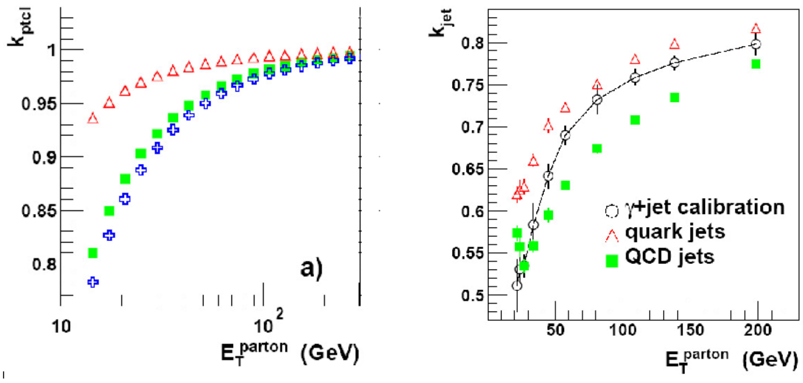

Figure 1.9: Left: distributions of the mean value of k_{ptcl} as a function of transverse parton energy for QCD di-jets (green square), for quark jets (open triangle) and for gluon jets (open crosses). Right: distributions of calibration coefficient obtained from γ+jets events (open circles) and their true value for generic QCD jet (full green circles) and quark jets (red triangles). Jets are reconstructed with $\Delta R = 0.5$ cone algorithm in the pseudo-rapidity region $|\eta| < 1.5$ [7].

values of k_{ptcl} are shown for generic QCD jets and for gluon and quark generated jets separately. The scale uncertainty due to the different fragmentation of gluon and quark generated jets is estimated by comparing the k_{ptcl} values obtained in the two cases. If $\Delta R = 0.5$ cone jets are considered the calibration coefficients differ by 5% for $E_T = 40$ GeV [19].
A second possibility to obtain calibration is to exploit kinematic constraints from real data such as the W mass in $W \to jj$ decays or the p_T balance in events where the jet is generated back-to-back with a well measured particle, either a Z decaying to leptons or a γ. In this note studies using γ+jet events are discussed.
ATLAS and CMS plan to use these events in different ways. CMS exploits the p_T balance constraint to obtain the calibration from calorimeter jet to parton jet while ATLAS plans to apply first the calibration to particle jet and than use the p_T balance constraint as a further step to correct to the parton jet energy scale. In the first phase of data taking the primary role of these events will be to help in understanding particle jet level calibration by comparing the data and Monte Carlo p_T balance distributions.

The selection of events in CMS requires a well isolated photon having a ϕ opening angle with the jet $\Delta\phi > 172^o$ [7, 19]. Events containing more than one jet with $E_T > 20$ GeV are rejected. The main background is given by QCD di-jet events where one jet is misidentified as a photon. Background is

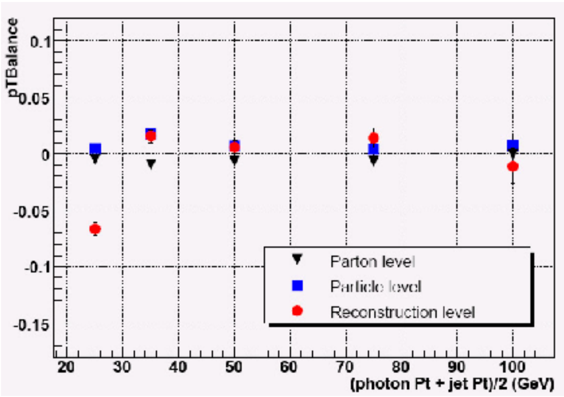

Figure 1.10: Distribution of $pTBalance = (p_T^{jet} - p_T^{\gamma})/p_T^{\gamma}$ as a function of $(p_T^{jet} + p_T^{\gamma})/2$ obtained by ATLAS on a sample of γ+jets events. The $pTBalance$ distribution is shown for calibrated calorimeter jets (full red circles), particle jets (blue triangles) and partons (black squares) [20]. Jet have been reconstructed with $\Delta R = 0.7$ cone algorithm.

suppressed well below the signal for $E_T^{\gamma} > 150$ GeV. The ratio $k_{jet} = p_T^{calo}/p_T^{\gamma}$ is calculated as a function of p_T^{γ} and defines the calibration coefficients. The complication given by the presence of initial state radiation that spoils the p_T balance constraint is partially overcome by defining, for each p_T^{γ}, the calibration coefficient to correspond to the most probable value of the p_T^{calo}/p_T^{γ} spectrum. The predicted values for the calibration coefficients and their true values ($k_{true} = p_T^{calo}/p_T^{parton}$) for quark jets and for jets from QCD background are shown in figure 1.10. At a transverse energy of 100 GeV a difference of about 10% is observed between QCD jets and quark jets. It should be noticed that this difference may be originated both by the different fragmentation spectrum of particles inside the jet and by the different out-of-cone losses. The p_T coverage of this channel after analysis cuts, indicates that, from a purely statistics evaluation, with $10 fb^{-1}$ a 1% statistical error is obtained up to a transverse energy of 800 GeV in the central region.
The event selection of ATLAS also starts with the requirement of a well isolated photons with $E_T^{\gamma} > 30$ GeV having an opening angle with respect to the highest p_T jet in the event of $\Delta\phi > 168^o$ [10, 20]. In order not to introduce a bias in the definition of the calibration coefficient due to the initial state radiation, the binning is done in bins of $(p_T^{\gamma} + p_T^{jet})/2$. The calibration coefficient in each bin, as for CMS, is defined as the most probable value of the p_T balance spectrum. Distributions of the p_T balance, defined as $(p_T^{jet} - p_T^{\gamma})/p_T^{\gamma}$, as a function of $(p_T^{jet} + p_T^{\gamma})/2$ are shown in figure 1.10. The three curves correspond to the p_T balance obtained using the jet calibrated to the particle

jet (as described in the previous section), the particle jet, and the parent parton. The balance obtained from particle jets and from calibrated jets agree within $\pm 2\%$ indicating that the particle level calibration, obtained on QCD di-jet events, may be applied also to different event topologies and different mixtures of partons. This result is somehow in disagreement with what is obtained by CMS (figure 1.10) where a large difference between quark and gluon jets is observed. It should be noticed, however, that the different cone size and the different correction for energy inside the cone makes it difficult to better understand the significance of this discrepancy. The particle level and parton level balance agree within $\pm 1\%$ indicating that underlying event contribution and the out-of-cone losses compensate each other to this level. Studies are ongoing to disentangle the two effects.

1.5 Energy Flow

Although the conceptual simplicity of calorimetric jets is a great asset for very early calorimeter understanding and calibration, an integration of the informations coming from the other detector components can provide a substantial improvement in both the measurement biases and the jet resolution.

In order to estimate the potential for improvement, one has to consider that 65% of the energy in an event is carried by charged particles (including the decays of unstable neutral particles into charged ones, the so called V^0's, like $K^0_S \to \pi^+\pi^-$ and $\Lambda^0 \to p\pi$), 25% by photons (including π^0 decays) and only 10% by long-lived neutral hadrons. This means that ideally, if all the photons were identified and corrected with specific calibrations and all the charged particles were measured by the tracking system, 90% of the energy could be better known. Additional improvement comes from particle identifications: not only electrons and muons would benefit from specific calibrations (since electrons loose most of their energy in the electromagnetic calorimeter and the muons deposit much less energy than hadrons in the calorimetric systems) but also V^0 recognition (since the measured invariant mass of the decay products can be replaced by the known mass of the "mother") and eventually the identification of the charged hadron as pion, kaon or proton (since all the particles, in jet, in first approximation are usually treated as pions, or even as massless particles, but at momenta of the same order of the particle mass this affects the energy measurement).

This ideal goal is made difficult by the unavoidable detector inefficiencies (e.g., the least energetic charged particles never reach the calorimeters due to the magnetic bending, so this part of the jet energy is unrecoverable) and by the identification ambiguities. Moreover, since the most important source of improvement is the replacement of the calorimetric measurement with the tracking information for charged hadrons, a critical factor is the ability of 1-to-1 association between tracks and calorimetric clusters, and this is limited by the coarseness of the calorimeter.

1.5.1 Energy Flow Algorithms in ATLAS

Inside the ATLAS collaboration, two different approaches to the use of the energy flow have been been studied. The first one [29] (approach A in the following) builds EnergyFlow objects from calorimeter towers and tracks and uses them as input objects for the jet reconstruction algorithm, while the second [30] (approach B) applies energy flow techniques on reconstructed jets. Both of them are at present somewhat limited by the *ad interim* solutions used inside ATLAS for the clustering. While at present the standard clustering for jets is done only in the η–ϕ space, the final clustering, which is under development, will make use of the complete η–ϕ-r segmentation of the ATLAS calorimetry, thus allowing for 3D clusters, more efficient in recognizing energy deposits belonging to a jet and less sensitive to noise.

The aim of the approach A is to define consistently topologically connected EnergyFlow objects. Each charged track seeds an EnergyFlow object. The tracks are then associated to calorimeter clusters both in the EM and in the HAD calorimeter extrapolating the track trajectory using the helix and making a matching in the η–ϕ space. The energy deposit expected for the particle (given its identification and its momentum measured by the tracker) is then subctracted from the calorimeter clusters. If the remaining energy in the cluster is within 1.28 σ_{noise} from zero, the cluster is removed from the cluster list. The remaining non–zero EM clusters seed EnergyFlow objects, the η–ϕ association is repeated and the expected energy deposits in the HAD clusters is subctracted. The remaining HAD clusters seed EnergyFlow objects.

Finally, EnergyFlow objects that are topologically connected (an EM cluster can be associated to more than one HAD cluster because of the bending of the magnetic field, for example) are grouped together in only one EnergyFlow object.

Approach B considers as input for the Energy Flow algorithm the already reconstructed jets. The idea is to identify (within a jet) clusters generated from charged hadrons, photons, electrons and finally neutral hadrons. To do this, a first iteration is performed on EM clusters. The central cell of those clusters that do not have a charged track pointing to them is chosen as a seed, and all the cells within $\Delta R = 0.0375$ are labelled as EMCL. Then an

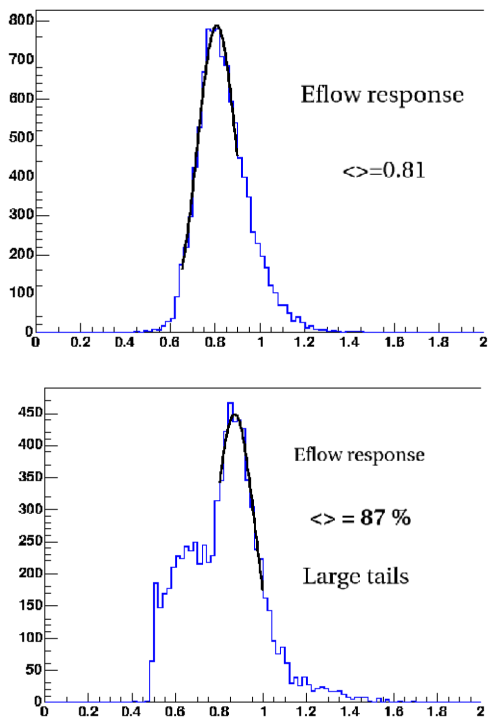

Figure 1.11: *The ratio between the reconstructed and reference energy is considered for events with 3 particles in the final state ($\gamma, n, \pi^{\pm}$). The shape of the distribution is degraded as they get close (on the left: $\Delta R > 0.1$, on the right: $\Delta R = 0.05$).*

iteration over the tracks is performed, and all the cells within $\Delta R = 0.0375$ from the track are labelled as CHRG. Finally, unassigned cells are labelled as NEUH. Ideally, EMCL should take into account photons, CHRG should account for charged pions, while NEUH should inlcude neutrons.

Has been already pointed out that the Energy Flow algorithms work at best with high granularity calorimeters and low multiplicity enviroment. If the subtraction of the expected energy is performed on an isolated cluster, one can expect an improvement on the resolution. But as soon as the clusters are not well separated, the subtraction of the expected value does not lead to an improvement of the resolution. This can be seen for example in fig. 1.11, where a "jet" composed by only three particles ($\gamma, n, \pi^{\pm}$) is considered. If the particles are far away in the η–ϕ space (left plot), the distribution of the measured energy is well shaped, but as soon as the particles become close (right figure), the Energy Flow response loose its regularity. Therefore, a refined 3D clustering algorithm is mandatory to improve the performances of the Energy Flow algorithms in ATLAS.

Fig. 1.12 shows the results of both the approaches discussed. Noise and pile–up are not included in the simulation. The left figure shows the current performances of approach A for 50 GeV jets. Two different contributions can be seen. The core of the distribution (whose $\sigma(E)/E$ is 7%) shows the performances where the track subtraction has worked, while in the broad peak, it did not work. The right figures shows the performances of approach B on jets with energy between 20 and 60 GeV. While the distribution is much more regular, the peak is broader ($\sigma(E)/E \simeq 12 - 13\%$) with respect to the core of the left plot. For comparison, the resolution quoted in the TDR for 50 GeV jets (from the standard calotimeter measurement) is 8%. The improvement of the clustering strategy could give an important improvement to the Energy Flow performances.

1.5.2 Energy Flow Algorithms in CMS

The improvement coming from the use of an Energy Flow technique is expected to be even more important for CMS than for ATLAS, due to their different detector designs: CMS has a more precise tracking system (thanks to the higher magnetic field and to the choice of using only pixel and microstrip silicon modules, while part of the ATLAS tracking system is constituted by the TRT, with coarser resolution), while the requirement of compactness makes its hadronic calorimeter less precise than the ATLAS counterpart. For this reason, a big effort is currently under way in CMS for the development of an optimal Energy Flow algorithm (actually called "Particle Flow", since particle identification plays a big role in it), with a large dedicated development group. This section presents only the first partial results towards this goal. Although these will be soon out of date and superseded by the complete algorithm, they show how much can be gained in CMS from the technique.

The simplest version [31] corrects the jet energy and direction after its reconstruction by the jet-finding algorithm (that uses the calorimetric deposits only).

The integration between Calorimeter and Tracking system measurements is performed by the EF algorithm through the following steps:

- Jets in the event are reconstructed by the calorimeter using an iterative cone algorithm. The jet object is defined by the collected energy and the direction.
- In the event all tracks with $P_T > 0.9$ GeV and $|\eta| < 2.4$ are reconstructed and selected at the vertex in a cone ΔR around jet direction. The cone is the same of the jet-finding algorithm.

- For each track the impact point on the ECAL inner surface is extracted and extrapolated to the HCAL one.
- The expected response of the calorimeter to each charged track is subtracted from the calorimetric cluster and track momentum is added.
- Other low P_T charged tracks, swept out of the jet cone definition by the magnetic field, are added to jet energy.

The algorithm performance has been tested comparing Montecarlo[4] and reconstructed jets, with and without EF applied. Di-jet events with P_T between 80 and 120 GeV/c were generated with PYTHIA and fully simulated and reconstructed inside the CMS software framework [34] [33]. Effects due to low luminosity ($L = 2 \times 10^{33} cm^{-2}s^{-1}$) pile-up have been included. The resolution and the reconstructed jet energy fraction are shown for jets generated with $|\eta| < 1.4$ in fig. 1.13. When the EF algorithm is applied, the reconstructed jet energy fraction for 40 GeV generated jets increases form 0.80 to 0.99 and the same fraction for 100 GeV jets increases from 0.85 to 1.00. The resolution improves by about 20-25% as a result of adding the out-of-cone tracks.

In the endcap region (figs. 1.14), jets with the same $\mathrm{E_T}$ as in the barrel are more energetic and, in addition, the tracking efficiency is smaller in the endcap than in the barrel. Therefore, the tracker information is not relevant in the endcap above 80-90 GeV and is less rewarding for lower $\mathrm{E_T}$ jets than in the barrel. Besides jets in the endcap are more affected by pile-up than in the barrel.

The performance of the EF algorithm has been tested also on events with a 120 GeV/c^2 X object decaying into light quarks with initial and final state radiation switched on. The X mass is reconstructed from the two leading jets that are within R = 0.5 of the direction of the primary partons. The ratio of the X mass reconstructed to the X mass generated for calorimetry jets and calorimeter-plus-tracker jets is shown in Fig. 1.15. The di-jet mass is restored with a systematic shift of about 1% and the resolution is improved by 10%. The ratio of the reconstructed to the generated X mass is 0.88 before corrections with tracks and 1.01 after corrections.

An improvement of the simple algorithm described above makes use of two cones with different size [32]: a smaller one for the jet-finding step and a larger one for the out-of-cone charged tracks recovery step. The idea of two different cones is suggested by the fact that neutral tracks release their

[4]Montecarlo jets are reconstructed implementing the same jet-finding algorithm than for reconstructed jet with tracks information from the MC truth

energy basically along the jet direction , since they are not deflected by the magnetic field. Therefore a small cone is sufficient to recover most of the neutral deposits in the calorimeter; the charged contribution to the jet energy is subsequently recovered by the tracker using a larger size cone. In this way, for the same amount of charged and neutral jet fragments recovered, the contamination by neutral deposit which do not belong to the jet (pile-up, underlying event, etc.) can be reduced.

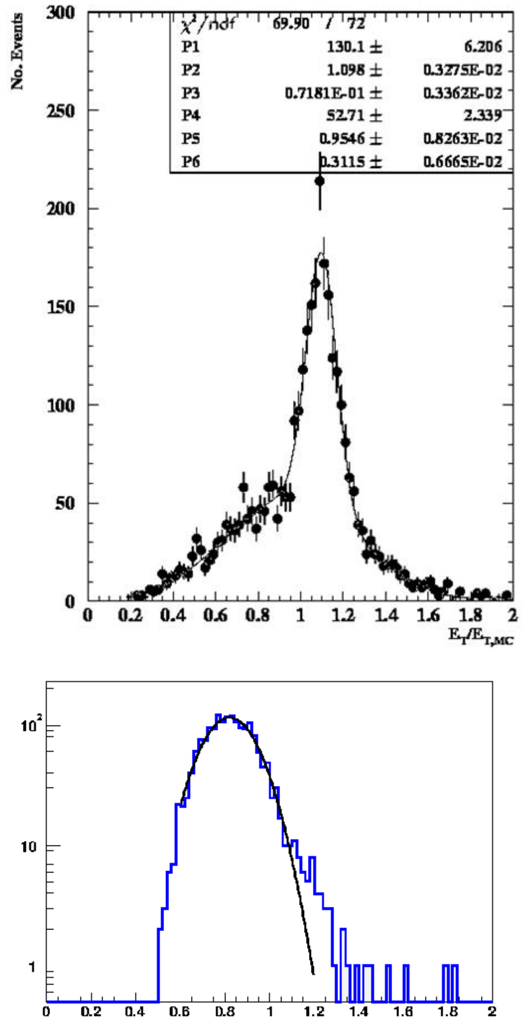

Figure 1.12: *On the left: the ratio between the reconstructed and the reference energy for the approach A on 50 GeV jets. The $\sigma(E)/E$ on the core of the distribution is 7%. On the right: The same for approach B for jets with energy between 20 and 60 GeV. The $\sigma(E)/E$ is 12–13%. As a reference, the TDR resolution for jets at 50 GeV is 8–9 %.*

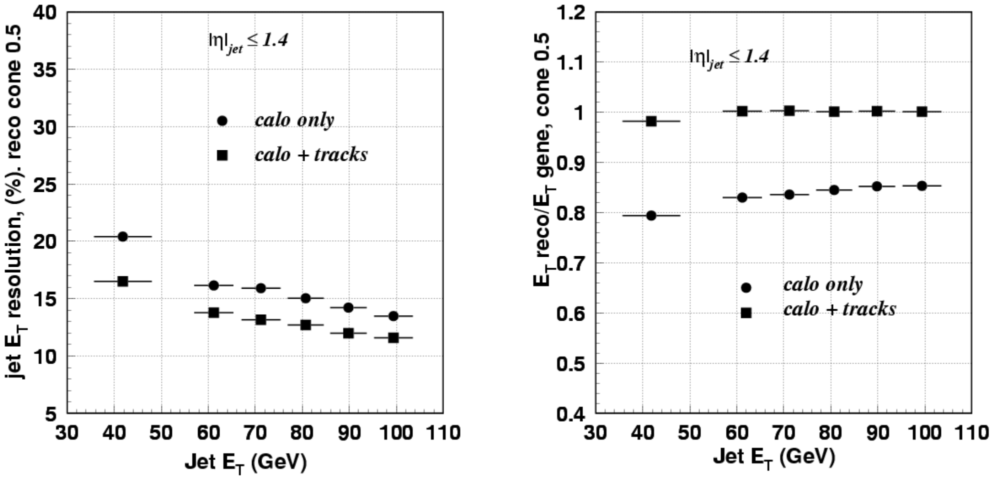

Figure 1.13: *Jet transverse energy resolution (left) and reconstructed jet transverse energy (right) as a function of the generated jet transverse energy. Jets with* $0 < |\eta| < 1.4$ *(barrel) from a sample with low luminosity pile-up; reconstruction with calorimeter only (close circles), subtraction procedure of expected responses using library of responses and out-of-cone tracks (close squares).*

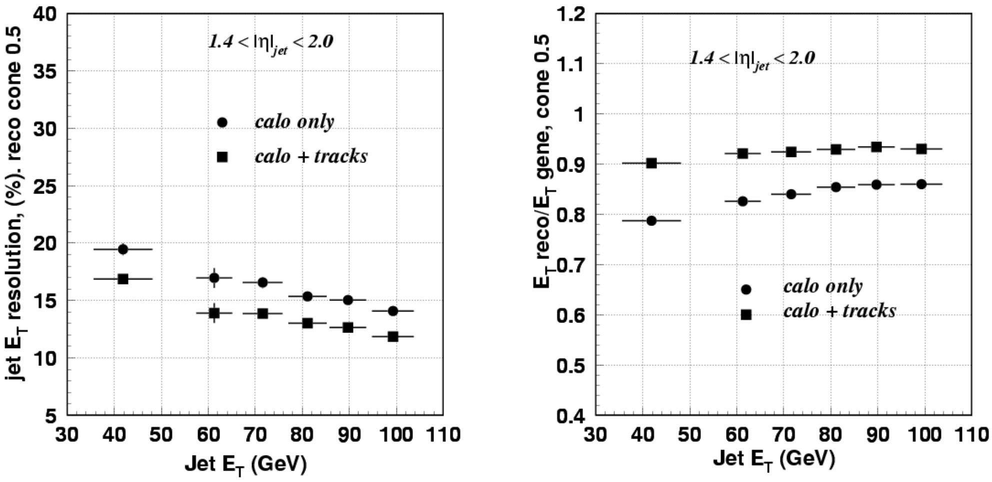

Figure 1.14: *Jet transverse energy resolution (left) and reconstructed jet transverse energy (right) as a function of the generated jet transverse energy. Jets with* $1.4 < |\eta| < 2.0$ *(endcap) from a sample with low luminosity pile-up; reconstruction with calorimeter only (close circles), subtraction procedure of expected responses using library of responses and out-of-cone tracks (close squares).*

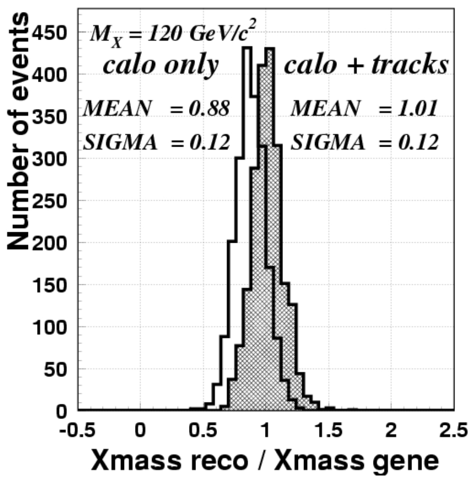

Figure 1.15: *Ratio of the reconstructed to the generated X mass with calorimeters only (empty histogram) and with calorimeter + tracks corrections (hatched histogram).*

Bibliography

[1] C. Cojocaru et al., Nucl.Instrum.Meth.A531:481-514,2004.

[2] D. Benedetti, S. Cucciarelli, A. Giammanco, J. Heyninck, J. D'Hondt, A. Schmidt, C. Weiser, *Study of jet clustering algorithms at the LHC*,

included in:

C. Buttar et al. *Les Houches Physics at TeV Colliders 2005, Standard Model and Higgs working group: Summary report*, hep-ph/0604120

[3] T.Sjöstrand, S.Mrenna, P.Skands, JHEP 05 (2006) 026.

[4] A. Pukhov et al., hep-ph/9908288.

[5] G.C. Blazey et al., hep-ex/0005012.

[6] CMS Collaboration, CMS HCAL TDR, CERN/LHCC 97-31, 1997.

[7] CMS Collaboration, CERN/LHCC 2006-001 CMS TDR 8.1.

[8] ATLAS Collaboration, CERN/LHCC/96-41; ATLAS Collaboration, CERN/LHCC/96-42.

[9] CMS Collaboration, Nucl. Instrum. and Methods A457(2001) 75-100.

[10] ATLAS Collaboration, CERN/LHCC/99-15.

[11] ATLAS Collaboration, Nucl. Instrum. and Methods A449(2000) 461-477;

[12] G.Blazey et al., hep-ex/0005012v2 10 May 2000.

[13] J.E.Huth et al.in "Proceedings of Research Directions For The Decade: Snowmass 1990", July, 1990, p. 134.

[14] ATLAS Collaboration, CERN/LHCC/96-40

[15] A.Heister et al., CMS NOTE 2006/036.

[16] HERWIG 6.5, G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M.H. Seymour and B.R. Webber, JHEP 0101 (2001) 010 [hep-ph/0011363]; hep-ph/0210213.

[17] A.Santocchia, *Optimization of Jet Reconstruction Settings and Parton-Level Correction for the* $t\bar{t}H$ *Channel*, CMS NOTE 2006/059

[18] A.Bhatti et al., hep-ex/0510047, 2005.

[19] O.Kodolova et al., CMS IN 2002/071.

[20] S.Jorgensen Roca, "γ+jet in-situ process for validation of the jet reconstruction with the ATLAS Detector", Universitat Autònoma de Barcelona, September 2006.

[21] ATLAS Calorimeter performance Technical Design Report, CERN/LHCC/96-40

[22] M.P. Casado and M. Cavalli Sforza, TILECAL-NO-75,1996

[23] S.Akhmadaliev *et al.*, Nucl. Inst. and Meth., A429, pag. 461, 2000

[24] Inner Detector Technical Design Report, CERN/LHCC 97–16 and 97–17

[25] The HCAL Technical Design Report, CERN/LHCC 97–31

[26] The ECAL Technical Design Report, CERN/LHCC 97–33

[27] V.V.Abramov *et al.*, Nucl. Inst. and Meth., A457, pag. 75, 2001

[28] The Tracker Technical Design Report, CERN/LHCC 98–6

[29] Talk given at 4th ATLAS Physics Workshop, Athens, 21–25 May 2003

[30] Talk given at ATLAS Software Workshop, 4th March 2004

[31] O.Kodolova et al., *Jet energy correction with charged particle tracks in CMS*, CMS NOTE 2004/015

[32] D.Spiga, A. Santocchia, L.Fano, *Double Cone Algorithm for Jet Energy Correction*, CMS IN 2005/013

[33] http://cmsdoc.cern.ch/orca/

[34] http://cmsdoc.cern.ch/cmsim/cmsim.html