Jets and jet algorithms

Victor Coco, Pierre-Antoine Delsart, Juan Rojo, Christian Sander Grégory Soyez

Jets are an important tool in hadronic physics and they will play a predominant role at the LHC. By defining jets as clusters of particles one aims at accessing, from the final-state particles, the underlying hard parton-level processes. Therefore jets are an essential tool for a variety of studies, such as top reconstruction, mass measurements, Higgs and new physics (NP) searches. Furthermore, they are instrumental for QCD studies, *e.g.* for inclusive jet measurements, which in turn constitute an important input for the determination of parton distribution functions. By clustering particles into jets, jet algorithms reduce complicated multiparticle events in simple final states with few jets. This procedure and the way particles are recombined together (*e.g.* the E- or P-scheme) is fundamentally non-unique.

In the following we will present recent progress in the description of jets, both from the phenomenological and the experimental points of view. In particular, we will focus on different aspects of the SISCone and anti-k_t jet algorithms. We will also describe jet finding strategies and jet reconstruction and calibration techniques being developed by the LHC experiments ATLAS, CMS and LHCb.

Finally, a recurring question in jet studies is what the best jet definition for a given physics analysis is. We will present a proposal of a characterization of jet-finding "quality" designed to be simple, robust, physical and reasonably representative of common analysis tasks.

1 The SISCone and anti-k_t jet algorithms

Author: Grégory Soyez

Two broad classes of jet definitions exist. The first one works by defining a distance between pairs of particles, performing successive recombinations of the pair of closest particles and stopping when all resulting objects are too far apart. Algorithms within this clustering class differ by the definition of the distance, frequent choices being $d_{ij}^2 = \min(k_{t,i}^2, k_{t,j}^2)(\Delta y_{ij}^2 + \Delta \phi_{ij}^2)$ for the k_t algorithm [1,2], and $d_{ij}^2 = (\Delta y_{ij}^2 + \Delta \phi_{ij}^2)$ for the Cambridge-Aachen algorithm [3,4].

Cone algorithms make up the second class, where jets are defined as dominant directions of energy flow. One introduces the concept of *stable cone* as a circle of fixed radius R in the $y - \phi$ plane such that the sum of all the momenta of the particles within the cone points in the same direction as the centre of the circle. Cone algorithms attempt to identify all the stable cones. Most implementations use a seeded approach to do so: starting from a given seed *e.g.*, a given direction for the centre of the cone, one computes the contents of the cone, takes the resulting momentum as a new direction and iterates until the cone is found stable. The set of seeds can be taken as the set of initial particles (sometimes over a p_t threshold) or as the midpoints between previously-found stable cones. As we shall see, this iterative method fails to identify *all* stable cones, leading to infrared (IR) or collinear unsafety in the perturbative computations.

Cone algorithms can be split into two sub-classes according to how they deal with the fact that stable cones may overlap. On the one hand, cone algorithms with split-merge identify the
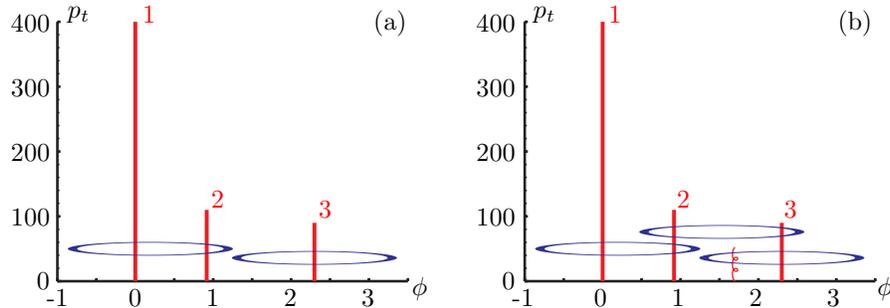
Fig. 1: Stable cones found by the midpoint algorithm for a 3-particle event (left) and for the same event with an additional infinitely soft gluon (right).

hardest overlapping pair of stable cones and merge (split) them if they share more (less) than a fraction f of the hardest cone. JetClu, midpoint and the ATLAS cone algorithms are typical representatives of that sub-class. On the other hand, cone algorithms with progressive removal start with the hardest unclustered particle, iterate from there until a stable cone is found and call it a jet. Its contents are removed and one starts again with the remaining particles. The CMS iterative cone is the typical example of this second sub-class, with the particular feature that hard jets are fully conical.

The Snowmass accords have established a series of requirements that any jet algorithm has to fulfill. These are basically that one can use the algorithm for theoretical computations, *e.g.* it gives finite perturbative results, as well as for experimental purposes, *e.g.* it runs fast enough and has small corrections from hadronisation and the underlying event.

We show in these proceedings that both the cone algorithms with split-merge and with progressive removal fail to give finite perturbative results. More precisely, we illustrate that midpoint suffers from IR unsafety and the iterative cone is collinear unsafe. We introduce SISCone and the anti-k_t algorithms as infrared- and collinear-safe solutions to those problems that do not spoil the experimental usability. We conclude by discussing the importance of using these new algorithms if we want to take full advantage of jet studies at the LHC.

1.1 SISCone as a replacement for the midpoint algorithm

Let us consider the 3-particle event displayed in Fig. 1(a). When clustered with the midpoint algorithm, 2 stable cones are found, leading to two jets: one with particles 1 and 2 and a second one with particle 3. If one adds to that hard event an infinitely soft gluon as shown in Fig. 1(b), a third stable cone is found and the three hard particles are clustered in a single jet. This change in the jet structure upon addition of soft particles, a phenomenon which happens with infinite probability in perturbative QCD, gives rise to divergences in the perturbative expansion and proves that the midpoint algorithm is IR unsafe[1]. Note also that the situation is even worse with JetClu or the ATLAS cone algorithms, where the IR unsafety is already present in events with 2 particles, *i.e.* one order earlier in the perturbative expansion.

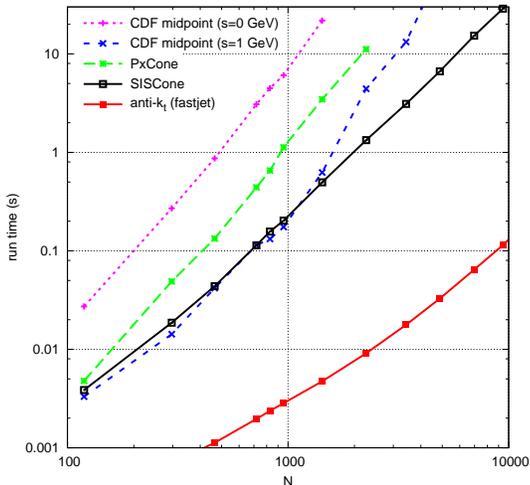[1]Note that when a seed threshold is used, the midpoint algorithm becomes collinear unsafe.

Fig. 2: Clustering time for SISCone compared to typical implementations of the midpoint and anti-k_t algorithms.

This problem arises from the fact that the seeded approach misses stable cones — here the one containing particles 2 and 3 in Fig. 1(a). The workaround to restore IR safety is thus to find a seedless method that provably identifies all stable cones. This is notoriously complex: a naive approach testing the stability of all subsets of particles [5] has a complexity of order $N \times 2^N$ for N particles which is much slower than the $\mathcal{O}(N^3)$ complexity of the midpoint algorithm, making this solution unusable for experimental purposes.

The solution [6] is to use the geometrical observation that any enclosure in the $y - \phi$ plane can be moved without changing its contents until it touches two points. Browsing all pairs of particles allows thus to enumerate all possible cones and to check their stability at an overall cost of $\mathcal{O}(N^3)$. Additional efforts to limit the amount of full stability tests to its minimum can even bring the final complexity to $\mathcal{O}(N^2 \log(N))$, *i.e.* faster than the midpoint algorithm. This has been implemented [6–9] in a C++ code named SISCone (Seedless Infrared Safe Cone). Fig. 2 illustrates the fact that in practice SISCone runs faster than the typical implementations of the midpoint algorithm without a seed threshold and at least as fast as when a 1 GeV seed threshold is used.

Therefore, SISCone is the first cone algorithm to satisfy the Snowmass requirements, that is to be at the same time IR and collinear safe, and to be fast enough to be used in experimental analysis.

1.2 Anti-k_t as a replacement for the iterative cone algorithm

As for the midpoint algorithm, we start by considering an event with three hard particles (see Fig. 3(a)). When clustered with the iterative cone, iteration starts with particle 2, one stable cone containing all particles is found, resulting in a 1-jet event. If we now split the hardest particle (2) into two collinear particles (2a and 2b) — a process that also happens with an infinite probability in perturbative QCD — as shown on Fig. 3(b), clustering with the iterative cone now starts with
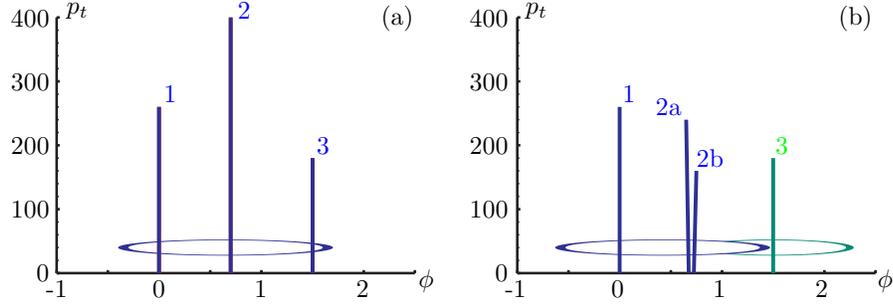
Fig. 3: Jets found by the iterative cone for a 3-particle event (left) and for the same event with a collinear splitting (right).

particle 1 which, after iteration, gives a first jet made of particle 1 plus the two collinear ones, then a second jet with particle 3. This example proves that the iterative cone algorithm is collinear unsafe.

Quite surprisingly, we can find a solution to that problem by coming back to the class of the recombination algorithms. The distance measures introduced earlier can be written as

$$d_{ij}^2 = \min(k_{t,i}^{2p}, k_{t,j}^{2p})(\Delta y_{ij}^2 + \Delta \phi_{ij}^2),$$

with $p = 1$ for the k_t algorithm and $p = 0$ for the Cambridge-Aachen algorithm. We can then consider a third case, the one for which $p = -1$ and call it the *anti-k_t algorithm* [10]. Obviously, this algorithm is IR and collinear safe. Furthermore, its implementation can benefit from the same geometrical observations that allowed for fast implementation of the k_t algorithm [8]. The anti-k_t algorithm thus runs at a speed similar to the one of the k_t algorithm, which certainly makes it usable for experimental purposes as seen on Fig. 2.

To understand the link between the anti-k_t algorithm and the iterative cone algorithm, we note from the definition of the anti-k_t distance that pairs involving a hard particle will be given small distances. This means that soft particles will be recombined with hard ones before recombining among themselves. As a result, the hard jets will have a circular boundary. This soft-resilience of the anti-k_t algorithm is exactly the hallmark of the iterative cone and it is in that respect that the anti-k_t can be seen as its IR and collinear safe replacement.

Fig. 4: Illustration of the regularity of the jets obtained with the anti-k_t algorithm.

To illustrate this property, we show in Fig. 4 the jets resulting from the clustering of an event made with a few hard particles and a large number of very soft ones uniformly dis-
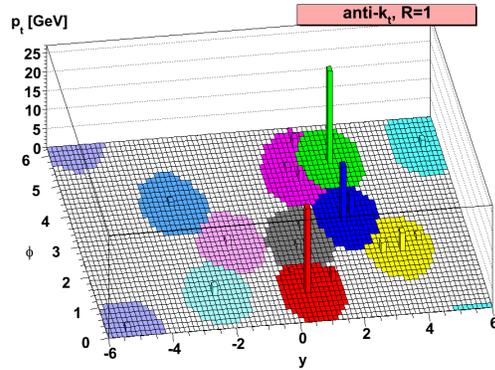
Observable	first miss cones at	Last meaningful order
Inclusive jet cross section	NNLO	NLO
$W/Z/H$ + 1 jet cross section	NNLO	NLO
3 jet cross section	NLO	LO (NLO in NLOJet)
$W/Z/H$ + 2 jet cross sect.	NLO	LO (NLO in MCFM)
jet masses in 3 jets	LO	**none** (LO in NLOJet)

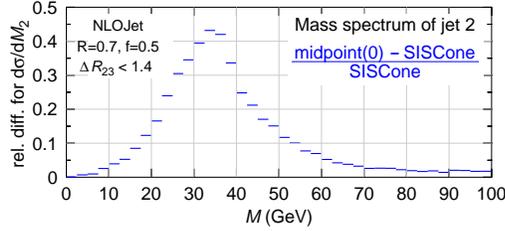Table 1: Perturbative level at which IR or collinear unsafety arises for various processes.

Fig. 5: Mass of the 2$^{\text{nd}}$ hardest jet in 3-jet events: relative difference between midpoint and SISCone. The 2$^{\text{nd}}$ and 3$^{\text{rd}}$ jets are imposed to be distant by at most $2R$.

tributed on a grid in the $y - \phi$ plane. It is clear that the hardest jets are perfectly circular and that, in general, the boundaries between the jets are regular.

1.3 Physical impact and discussion

As we have seen, the seeded approach to stable cone search suffers from problems with respect to perturbative QCD expansion: the algorithms with split-merge are IR unsafe, while the iterative cone (with progressive removal) is collinear unsafe. We have introduced SISCone as a natural replacement of the cone algorithms with split-merge like midpoint, and the anti-k_t algorithm as a candidate to replace the iterative cone. These new algorithms are both IR and collinear safe.

The question one might ask is to what extent these IR and collinear safety issues are important in real measurements. Since the unsafety arises when one has 3 particles in a common vicinity, it becomes important at the order α_s^4 or $\alpha_{EW}\alpha_s^3$ of the perturbative series.

Table 1 summarises for different physical processes, the order at which seeded algorithms like midpoint of the iterative cone stop to be valid. The main message we can get from that table is that, if we do not want theoretical efforts in precise QCD computations to be done in vain, the resort of an IR and collinear safe algorithm like SISCone and the anti-k_t is fundamental.

To illustrate the argument more quantitatively, Fig. 5 shows the relative difference, expected to be present at the LO of perturbative QCD, between SISCone and midpoint for the mass of the second hardest jet in 3-jet events. Differences reaching up to 40% are observed, proving that an IR and collinear safe algorithm is mandatory. The situation is even worse with JetClu or the ATLAS cone algorithm. As the infrared-unsafety problem becomes apparent at the order α_s^3 or $\alpha_{EW}\alpha_s^2$, *i.e.* one order earlier than with midpoint.

2 Quality measures for jet finding at the LHC

Author: Juan Rojo

A recurring question in jet studies is what the best jet definition for a given physics analysis is. In this contribution we propose a characterization of jet-finding "quality" designed to be simple, robust, physical and reasonably representative of common analysis tasks.

For this purpose, we require a source of quarks and gluons with well-defined energies. We will obtain these from Monte Carlo production and decay of fictitious narrow Z' and H bosons, with $Z' \to q\bar{q}$ and $H \to gg$ generated with Pythia 6.5 [11] with di-jet invariant masses ranging from 100 GeV to 4 TeV. For each generated event we will cluster the event into jets with about 50 different jet definitions, where a jet definition, JD, consists of the jet algorithm and the associated parameters, like the radius R [12]. The radius R will be varied between 0.3 and 1.5. For each event, we determine the invariant mass of the sum of the two hardest jets. The distribution of invariant masses should then have a peak near the heavy boson mass. We will take the sharpness of that peak to be indicative of the quality of each jet definition.

The infrared- and collinear-safe (IRC) safe jet algorithms under scrutiny are the longitudinally invariant inclusive k_t algorithm [1,2,13], the Cambridge-Aachen (C/A) algorithm [3,4], the anti-k_t algorithm [10], SISCone [6] as well as C/A with filtering. The latter is C/A supplemented with a filtering procedure [14] in which, subsequent to the jet finding, each jet is unclustered down to subjets at angular scale $x_{\text{filt}} R$ and one retains only the n_{filt} hardest of the subjets. We use $x_{\text{filt}} = 0.5$ and $n_{\text{filt}} = 2$. All the jet algorithms have been used in the implementations and/or plug-ins of the `FastJet` package [8], version 2.3, with the exception of C/A with filtering, which will be made public in a forthcoming `FastJet` release.

This contribution summarizes work [15] in collaboration with M. Cacciari, G. Salam and G. Soyez, initiated in the context of the "Les Houches Physics at TeV colliders 2007" workshop [12].

2.1 Quality measures and effective luminosity ratio

As described in detail in [15], the merit of the jet finding is quantified by two quality measures:

1. $Q_{f=z}^{w}$: the width of the smallest (reconstructed) mass window that contains a fraction $f = z$ of the generated massive objects,

$$f \equiv \left(\frac{\text{\# reco. massive objects in window of width } w}{\text{Total \# generated massive objects}} \right) = z \,. \tag{1}$$

2. $Q_{w=x\sqrt{M}}^{1/f}$: to compute this quality measure, we take a window of fixed width w and slide it over the mass distribution so as to maximise its contents. Then the figure of merit is given by

$$Q_{w=x\sqrt{M}}^{1/f} \equiv \left(\frac{\text{Max \# reco. massive objects in window of width } w = x\sqrt{M}}{\text{Total \# generated massive objects}} \right)^{-1} , \tag{2}$$

It is clear from its definitions that the smaller the quality measures, the better the corresponding jet definition. An illustrative example of these two measures is shown in Fig. 6. We observe that the quality measures quantify the intuitive assessment of the goodness of jet finding, represented by the sharpness of the reconstructed invariant mass peak. Note that in our approach, any matching to non-physical quantities like Monte Carlo partons is deliberately avoided.
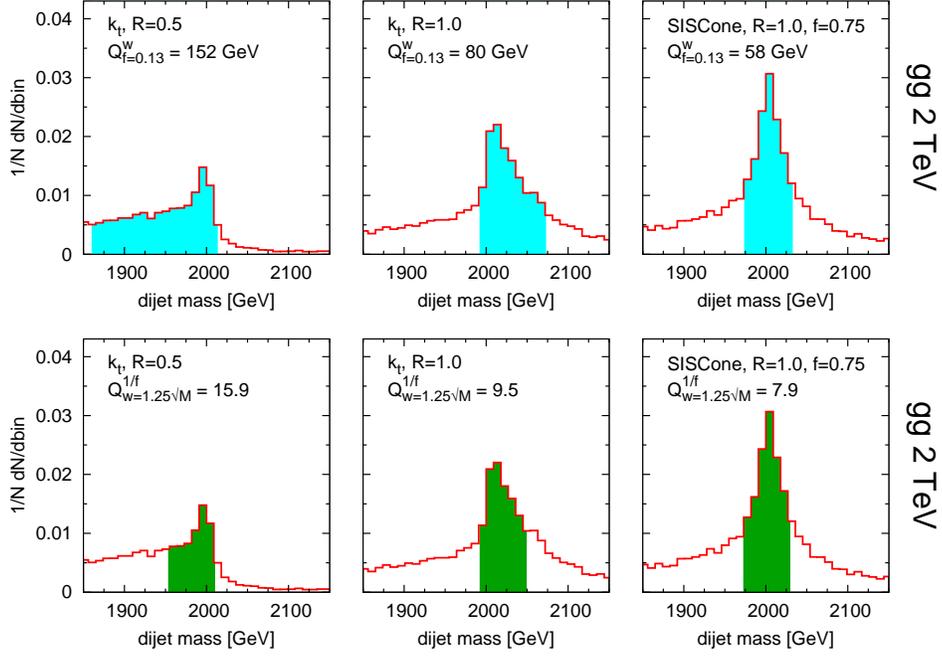

Fig. 6: Di-jet invariant mass distributions for the gg case at $M = 2$ TeV, comparing three jet definitions for each process. The shaded bands indicate the region used when obtaining the two different quality measures.

These quality measures can be mapped to the corresponding variation of integrated luminosity needed to maintain constant signal significance. As we have seen, a larger quality measure indicates a worse jet definition. This in turn implies that a larger luminosity will be needed to obtain a given significance. It is convenient to express this in terms of an effective luminosity ratio,

$$\rho_{\mathcal{L}}(\mathrm{JD}_2/\mathrm{JD}_1) \equiv \frac{\mathcal{L}(\text{needed with } \mathrm{JD}_2)}{\mathcal{L}(\text{needed with } \mathrm{JD}_1)} = \left[\frac{\Sigma\left(\mathrm{JD}_1\right)}{\Sigma\left(\mathrm{JD}_2\right)}\right]^2 . \tag{3}$$

with the signal significance defined in the usual way $\Sigma\left(\mathrm{JD}\right) \equiv N_{\mathrm{signal}}^{\mathrm{JD}}/\sqrt{N_{\mathrm{bkgd}}^{\mathrm{JD}}}$. Given a certain signal significance with JD_1, $\rho_{\mathcal{L}}(\mathrm{JD}_2/\mathrm{JD}_1)$ indicates the factor more luminosity needed to obtain the same significance with JD_2. For example, the expression for ρ_L in terms of the first quality measure is

$$\rho_{\mathcal{L}}(\mathrm{JD}_2/\mathrm{JD}_1) = \frac{Q_{f=z}^{w}\left(\mathrm{JD}_2\right)}{Q_{f=z}^{w}\left(\mathrm{JD}_1\right)} . \tag{4}$$

A non-trivial check of the robustness of our analysis is that the luminosity ratios obtained with the two different quality measures are roughly consistent with each other.
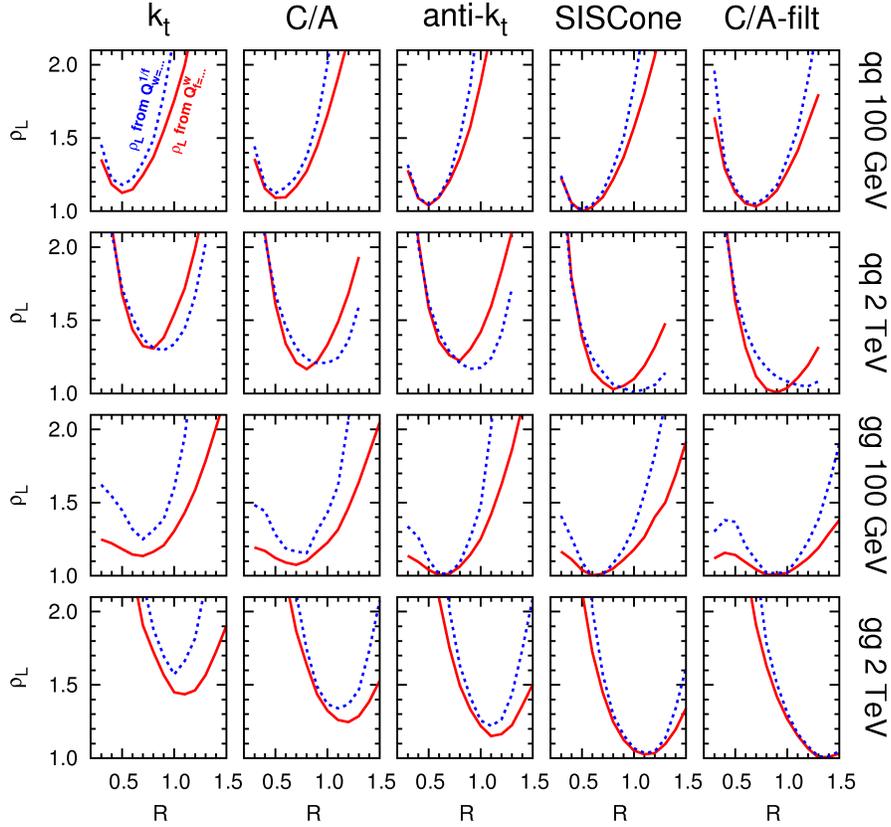
2.2 Results

Fig. 7: The effective luminosity ratio, Eq. 3, for quark and gluon jets at 100 GeV and 2 TeV, for all algorithms studied. The two curves in each plot correspond to the value of $\rho_\mathcal{L}$ computed from the respective quality measure. For each process, $\rho_\mathcal{L}$ is normalized to the corresponding optimal jet definition.

Now we present selected results for the effective luminosity ratio for the different cases considered. We show in Fig. 7 a summary of the performance of the various jet definitions studied, for quark and gluon jets at 100 GeV and 2 TeV, without pile-up (PU). First of all, we observe a strong dependence of ρ_L with respect to R, as well as sizable differences between jet algorithms. SISCone and C/A-filt turn out to be the optimal jet algorithms in all studied processes. They achieve limited sensitivity to the Underlying Event (UE) while maintaining their perturbative reach. The optimal value of R grows with the scale of the process, specially for gluon jets, reflecting the interplay between perturbative and non-perturbative effects [16]. Our studies imply that at the TeV scale, rather large values of $R \sim 1$ are required to obtain optimal
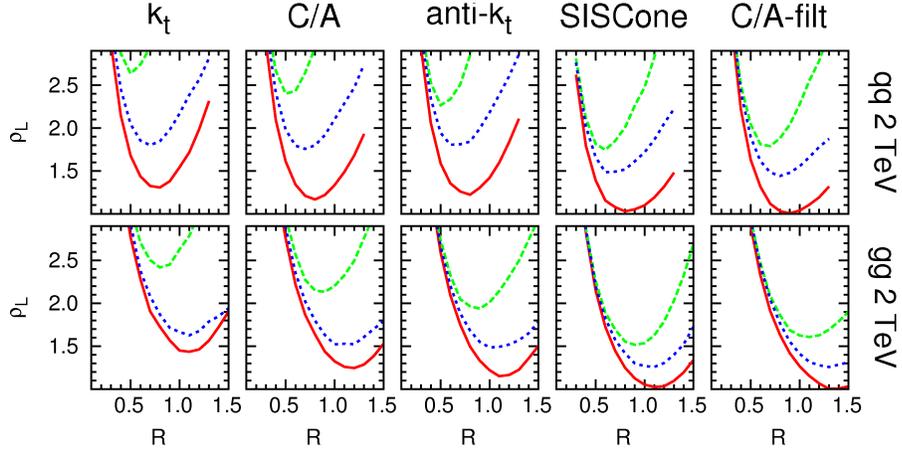
Fig. 8: The effective luminosity ratio, for quark and gluon jets at 100 GeV and 2 TeV, for all five algorithms studied. The red solid lines correspond to the no-PU case, the green dashed lines to the high luminosity PU case while the blue dotted curves correspond to high luminosity with PU subtracted as explained in the text.

resolution. LHC experiments, on the other hand, plan to use smaller radii in general, see for example Ref. [17].

From Fig. 7 one can determine how much more luminosity will be required with a less favoured jet definition compared with the optimal one. For example, we see that for the gg case at 2 TeV, if the k_t algorithm is used instead of the optimal one (SISCone), then 50% more luminosity will be required to achieve the same signal significance even at the respective optimal values of R.

These results are robust against high-luminosity PU [15] once PU is subtracted using the `FastJet` area method [18, 19], as can be seen in Fig. 8. This has the important consequence that for a given process, a single jet definition could be used at the LHC regardless of the machine luminosity.

As a practical application of our studies, one can consider the impact of less favoured jet definitions in LHC searches with similar signatures. For example, let us consider a particular scenario in which a di-jet invariant mass distribution is reconstructed and let us assume that the jet clustering is performed with a jet definition, JD_2, whose quality is far from the optimal one, JD_1, so that the effective luminosity ratio is large, say $\rho_{\mathcal{L}} \sim 2$. The net effect of the choice of such non-optimal jet definition for the kinematical reconstruction can be summarized schematically in Fig. 9: the use of JD_1 rather than JD_2 would lead to a discovery signal with approximately only half of the machine running time required with the original jet definition.
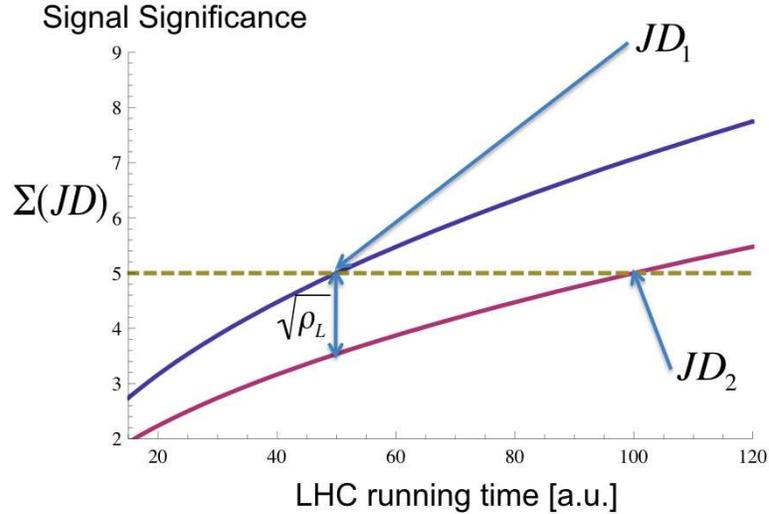
Fig. 9: Example of how optimizing the jet definition might lead to discoveries in less machine running time, compared to the non-optimal one.

2.3 Conclusions

Summarizing, we have proposed a technique to quantify the performance of jet algorithms for kinematic reconstructions at the LHC. To allow for more detailed studies of the results of Ref. [15], an interactive webpage has been created at `http://quality.fastjet.fr`, which allows the user to test the effects of changing and modifying various jet definitions and other inputs like PU luminosity for the process under scrutiny.

3 Performance of jet reconstruction at CMS

Author: Christian Sander (on behalf of the CMS Collaboration)

Almost every process of interest at the LHC contains quarks or gluons in the final state. The partons can not be observed directly, but fragment into stable hadrons, which can be detected in the tracking and calorimeter systems. Calorimeter jets are expected to yield a good description of both the parton-level and the hadron showers emerging from the hard interaction. For Monte Carlo (MC) events, the hadron-level is defined by applying the same clustering algorithms, which are typically formulated to accept any set of four-vectors as input to all stable particles from the MC truth record ("GenJets"). Hadron-level is also referred to as "particle-level", and jet energy scale corrections based on MC and later on data-driven methods are derived to correct back to this detector independent level. Calorimeter jets are reconstructed using energy deposits in calorimeter towers ("CaloTowers") as inputs: they are composed of one or more hadronic calorimeter (HCAL) cells and corresponding electromagnetic calorimeter (ECAL) crystals.

The studies presented in what follows are based on QCD di-jet and $t\bar{t}$ MC samples without pile-up. It is often necessary to associate CaloJets with GenJets in these samples to probe how well the calorimeter-level reconstruction represents the hadron-level of the process. This

association is based on spatial separation in the η-ϕ-space between the two jet axes by requiring

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$

to be less than a certain value. Besides good correspondence to the parton-level and hadron-level, a successful jet algorithm should fulfill two important requirements. Firstly, it should be collinear-safe, such that the outcome remains unchanged if *e.g.* the energy carried by a single particle is instead distributed among two collinear particles. Collinear safety is typically endangered if the jet finding is based on energetic seeds and a threshold is applied to these seeds. Secondly, it should be infrared-safe, such that the result of the jet finding is stable against the addition of soft particles. Jet algorithms which don't comply with either or both of these requirements yield ambiguous results and lead to unnecessary uncertainties when applied to calculations in perturbative theory. The performance of the following four jet clustering algorithms is discussed:

- The Iterative Cone algorithm is a simple seeded cone-based algorithm employed by CMS online in the High Level Trigger (HLT). It has a short and predictable execution time, but is neither collinear- nor infrared-safe.
- The Midpoint Cone [5] algorithm is similar to the Iterative Cone, but infrared-safety is addressed by considering the midpoints between each pair of close (proto-)jets as additional seeds. Despite its improvements to the cone-based clustering procedure, the algorithm has been shown not to be infrared-safe. This algorithm is no longer supported by CMS.
- SISCone [6] is the "Seedless Infrared-Safe Cone" jet algorithm. It is collinear- and infrared-safe to all orders of perturbative QCD (pQCD) and demands only slightly higher execution time compared to the Midpoint Cone algorithm.
- fast-k_T [8] is a recent implementation of the k_T algorithm [1] which is also collinear- and infrared-safe. It has a dramatically reduced execution time with respect to previous implementations of the k_T algorithm.

The "E-Scheme" is used for all algorithms as the recombination scheme: the energy and momentum of a jet are defined as the sums of energies and momenta of its constituents. The execution time of the fast-k_T algorithm is comparable to the Iterative Cone algorithm without the discussed deficiencies of the latter. The SISCone algorithm requires more CPU resources compared to the Midpoint Cone algorithm. The time spent for the jet reconstruction (0.02 s) of each event however is small compared to the total event reconstruction time (10 s): the particular jet algorithm choice does not impact the overall CPU requirements.

3.1 Summary of Jet Performance Study

The performance of the CMS calorimeters is known to be different in the barrel, endcaps and forward regions. Here we focus on the relative performance between different algorithms and radius parameter choices currently supported for CMS analysis. Only distributions for the barrel region are therefore shown. Further details can be found in [20].

The jet matching efficiency is defined as the ratio of the number of particle jets matched to a calorimeter jet within $\Delta R < 0.5$ to the total number of particle jets. It represents a meaningful measure of the reconstruction efficiency of each jet algorithm, but is strongly correlated

to the position resolution and therefore depends on the ΔR cut and the jet size parameter. However, relative comparisons between different algorithms using equivalent size parameters remain instructive. The matching efficiencies for small (left) and large (right) radius parameters as a function of the MC truth p_T^{gen} are shown in Fig. 10. The efficiencies of jets reconstructed with the fast-k_T and SISCone algorithms indicate better performance than jets reconstructed with the MidpointCone and Iterative Cone algorithms.

For the jet response, $R_{\text{jet}} = p_T/p_T^{\text{gen}}$, very good agreement between the individual algorithms is found for all regions of the detector, indicating good correspondence between the values of D for the fast-k_T algorithm and R for cone algorithms which are being compared [20].

The η resolutions for jets in the barrel region are shown as a function of p_T^{gen} in Fig. 11. Good agreement is found among all algorithms with comparable radius parameter, with marginal differences at low p_T^{gen}. Jets reconstructed with larger radius parameters yield slightly worse resolution. Note that the position of the primary vertex is assumed to be at $z = 0$, which dilutes the η resolution with respect to taking the correct position measured with the tracking detectors into account. The ϕ resolutions can be found in [20].

Fig. 12 shows the jet energy resolutions derived from MC truth for jets in the barrel region. Jets reconstructed with fast-k_T show slightly worse resolution at low p_T^{gen}, while no significant impact of the radius parameter choice is observed. The typical jet energy resolution at high p_T, 100 GeV or 1 TeV, is $\sim 14\%$ and $\sim 7\%$ respectively, with no significant dependence on the jet clustering algorithm.

The jet reconstruction performance in $t\bar{t}$ events is studied by selecting events with one ("lepton+jets") or zero ("alljets") electron(s) or muon(s) in the final state from a $t\bar{t}$ sample with no additional jets ("$t\bar{t}$+0 jets"). $t \rightarrow bq\bar{q}'$ and $\bar{t} \rightarrow \bar{b}\bar{q}q'$ decays are identified on particle level and only events are considered for which all three decay products of one or both $t(\bar{t})$ decay(s) can be uniquely matched to reconstructed calorimeter jets. The efficiency to select these decays indicates the performance of the respective jet algorithm in a busy multi-jet environment and its ability to correctly resolve the topology of the underlying process. The fast-k_T algorithm is hereby found to fully resolve hadronic $t(\bar{t})$ decays on calorimeter level more efficiently than any cone-based algorithm. For the selected events, the invariant two-jet (W boson) and three-jet (top quark) masses are compared on particle-level, calorimeter-level, corrected calorimeter-level, and corrected calorimeter-level with additional flavor-dependent corrections applied. The m_W and m_t distributions obtained for all correction levels are shown in Fig. 13 for jets reconstructed with fast-k_T $D = 0.4$. From the width of the obtained invariant mass distributions one can see that the impact of detector effects on the mass resolution are stronger than the algorithmic differences. A full comparison of the widths of the reconstructed m_W and m_t distributions can be found in [20].

3.2 Conclusion

The performance comparisons presented include jet energy response, position resolutions, energy resolutions and efficiencies in QCD di-jet samples. We find similar performance at the calorimeter level between algorithms with similar size parameter. The impact of detector effects appears to be more pronounced than the algorithmic differences studied here. The SISCone algorithm performs as well as or better than the Midpoint Cone, while known to be preferred

theoretically. Therefore it was decided to adopt SISCone as the default cone-based jet algorithm and consequently to include it in the reconstruction in the standard event processing at CMS.

The fast-k_T algorithm is infrared- and collinear safe to all orders of pQCD as well and complementary to the cone-based algorithms. The execution time of fast-k_T is dramatically reduced with respect to earlier implementations and it is therefore well suited for the high multiplicity environment of LHC pp collisions. We find that it performs as good or better than any other compared algorithm and strongly encourage its use as an alternative to SISCone.
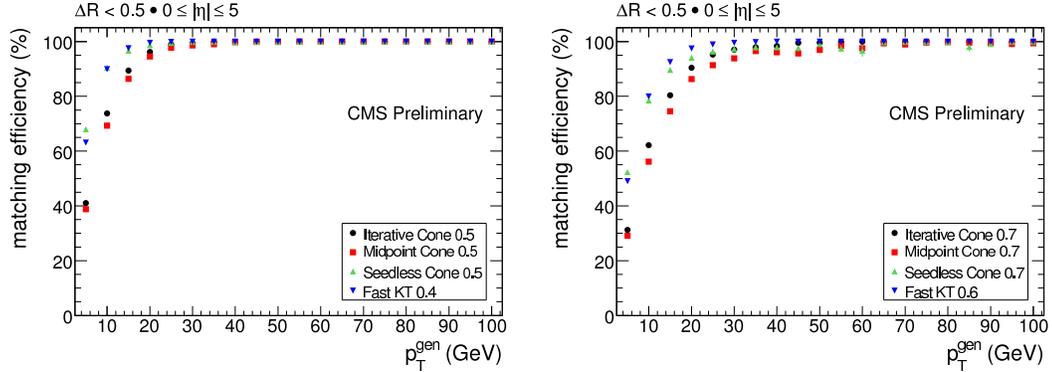

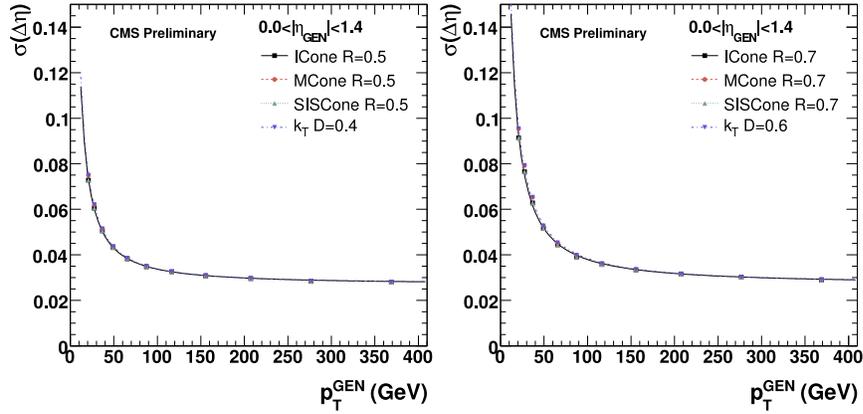Fig. 10: Matching Efficiency versus p_T^{gen} for $R = 0.5/D = 0.4$ (left) and $R = 0.7/D = 0.6$ (right) jets.

Fig. 11: The jet η resolutions as a function of p_T^{gen}, averaged over the Barrel region, for jets clustered with smaller (left) and larger (right) size parameters. The resolutions are derived using MC truth information.

4   Jet finding strategies in ATLAS

Author: Pierre-Antoine Delsart (on behalf of the ATLAS Collaboration)

ATLAS is a general purpose experiment at the Large Hadron Collider (LHC) [21]. Its calorimetry system, the principal tool for hadronic jet measurements, is described in detail in [21], chapter 5. Some key features of this calorimeter relevant to jet finding are its wide
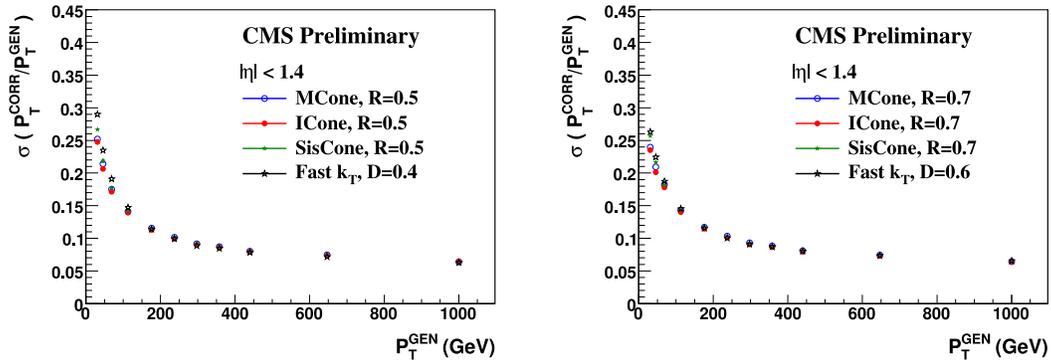
Fig. 12: Jet energy resolution derived from MC truth for Midpoint Cone, Iterative Cone, SISCone and fast-k_T with $R = 0.5/D = 0.4$ (left) and $R = 0.7/D = 0.6$ (right) in the barrel region ($|\eta| < 1.4$).
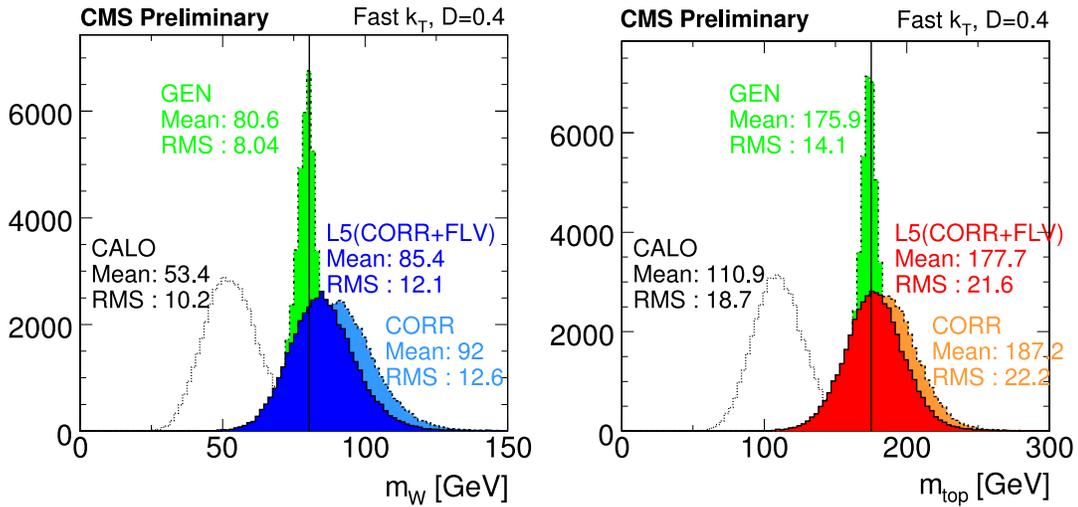

Fig. 13: m_W and m_t distributions for hadronic top decays reconstructed with the fast-k_T algorithm, $D = 0.4$. Distributions are shown for particle-level jets (GEN), calorimeter jets (CALO), calorimeter jets corrected with "MC-Jet" corrections (CORR), and corrected calorimeter jets with an additional flavour correction applied (L5). Only jets with uncorrected $p_T \geq 15$ GeV and $|\eta| \leq 5$ are considered. The generated W boson (80.42 GeV) and top quark (175 GeV) masses are indicated by the black vertical lines.

acceptance (up to $|\eta| = 4.9$ in the Forward Calorimeter) and a fine granularity (including up to 7 longitudinal segmentations). On the other hand, the calorimeter is non-compensating (ratio $1.3 < e/\pi < 1.6$ depending on the specific sub-calorimeter) and this causes the major source of uncertainty in energy measurements because of the large fluctuations of the electromagnetic component of hadronic showers.

The other main experimental challenge will come from the LHC environment : a very large phase space for underlying event, multiple interactions per bunch crossing (23 at full luminosity). Out-of-time pile-up is also expected because of the slow response of the liquid argon calorimeter which will integrate several events before and after a given interaction.

In order to take up these challenges the ATLAS collaboration chose to adopt a flexible approach. In particular, the collaboration is studying two calibration strategies, several in-situ and data-based correction methods, and has designed a software able to cope with any jet algorithm used in physics analysis.

4.1 Jet reconstruction and calibration

The ATLAS jet-related software is designed to allow any input to jet finding algorithms, provided the input is a set of valid four-momenta. This allows to run exactly the same jet finders on Monte Carlo truth simulated particles, real signal, tracks, etc. Two different calorimeter signal definitions are considered as input signal for jet finding:

- Calorimeter towers : all cells in the same projective direction (defined by a grid in the (η, ϕ) plane) are grouped into a tower. The four-momentum is formed by the sum of the cells energies, possibly including a geometrical weight for cells larger than the tower grid size, and the direction of the tower.
- Topological clusters ("TopoClusters"). Cells are clustered together in the 3 dimensions of the calorimeter according to a nearest neighbour algorithm [21] which intrinsically performs a noise suppression.

Besides different types of input signal, ATLAS considers two approaches for the jet calibration.

Global hadronic calibration. Jets are built from raw calorimeter signal (towers or clusters), then a set of correction factors (weights) are applied to the energy of the constituting cells. The weights depend on the characteristics of the cells, in particular its energy density and its location in the calorimeter. They are extracted from a fit to simulated di-jet events. With this method, all calibration corrections are included in a single set of weights, hence its name "global" [22].

Local hadronic calibration. This second method is an attempt to have a finer, better understood calibration method for jets. It relies on hadronic calibration of topological clusters [23] : jets are built from these calibrated input signals. Then a jet energy scale correction remains to be applied. This approach is more complex but allows to decouple different corrections (non-compensation, dead material losses, energy scale) and is thus very promising.

An illustration of performance for both calibration methods is shown in Fig. 14; a detailed discussion of these performances can be found in Ref. [24];
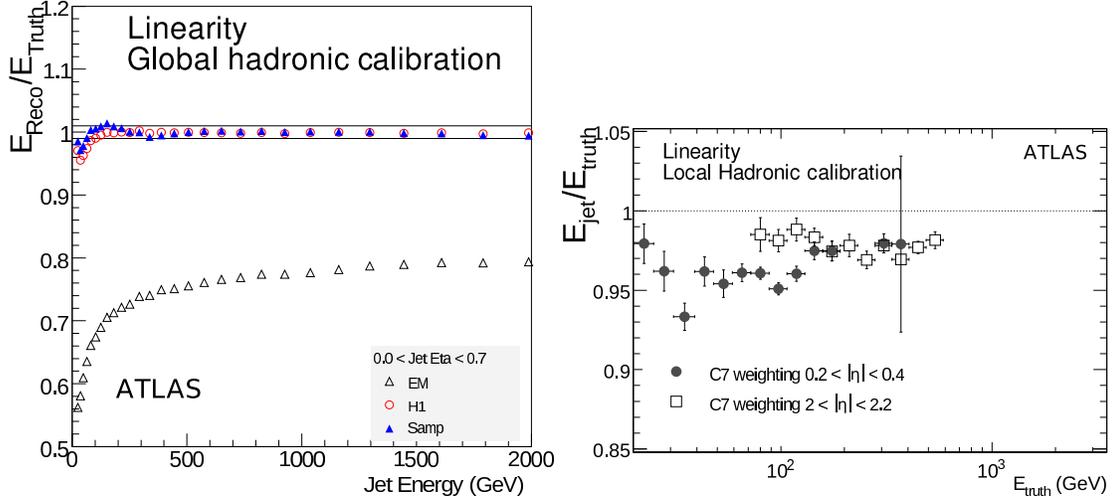
Fig. 14: Left : linearity of jet reconstruction with global calibration (QCD di-jet sample). Red and blue marks correspond to 2 alternative global calibration methods. Right: linearity with local calibration (before energy scale correction).

4.2 Data driven corrections

In parallel to the base-line calibrations described above, ATLAS aims to reach a precise energy scale measurement using experimental data directly. Several methods are studied:

- Momentum measurement from the tracker (P) compared to energy deposition in calorimeter (E) allows to validate the energy scale for charged pions in minimum bias events [25], studying the E/P ratio.
- QCD di-jet events can be used to uniformize the response of the calorimeter in η and ϕ [26].
- Z+jets or γ+jets events will be used assuming an excellent calibration of electromagnetic objects : applying P_T balance or missing E_T projection techniques will allow to retrieve a correct jet energy scale [26].
- With QCD multi-jet events it is possible to correct high-p_T jets against several lower p_T jets whose energy scale is better known thanks to previous methods [26].

In some analyses, *in-situ* methods using constraints coming from the mass of the W boson will be applied to control even better the jet energy. This is typically the case in top physics analyses where different methods are under study [27].

4.3 Jets algorithms, other jets studies

Several different jet algorithms are available for physics analysis. Two families of such algorithms are reconstructed by default :

- ATLAS iterative cone algorithm (described in detail in [22]), with cone radii 0.4 and 0.7;
- Kt clustering algorithm [22], setting the size parameter D to 0.3 and 0.6 .

These algorithms are officially supported and used in calibration studies. Variations of these algorithms with different jet sizes and clustering parameters can easily be configured, as appropriate in the context of a given physics analysis. In addition, other algorithms like the midpoint cone algorithm [28], the seedless infrared safe cone algorithm SISCone [6] and all flavours of recursive recombination algorithms provided in the FastJet [8] library, and the "optimal jet finder" described in [29], are available within the standard ATLAS software framework.

Various other jet-related studies are on-going in the ATLAS collaboration in order to understand better and improve jet reconstruction:

- Associating reconstructed tracks with calorimeter signals allows to obtain efficient jet energy corrections. Moreover, vertex information can help in rejecting jets coming from pile-up [24].
- Different studies are on-going in order to understand precisely the effects of pile-up, in particular in the liquid argon calorimeter.
- Jets sub-structure studies such as the use of the "y-scale" given by k_t jet algorithms [30]

4.4 Conclusion

We gave a brief overview of the recent work of the ATLAS collaboration related to jets reconstruction (details in [31]). In order to deal with the great experimental challenges and to achieve an excellent measurement of hadronic jets, the collaboration has adopted a flexible approach including two main strategies for jet calibration. The collaboration is also preparing several data-based and in-situ techniques to correct and control the jet energy scale and resolution at the precision required by physics analysis as well as conducting several studies to ensure the understanding of the detector response to hadronic jets is optimal.

5 b-jets at LHCb

Author: Victor Coco (on behalf of the LHCb Collaboration)

LHCb [32] is an LHC experiment dedicated to precise measurements of CP violation and rare B-meson decays. We show that its specifications are of interest for reconstruction and identification of b-jets as well. The LHCb detector is a one-arm spectrometer. It covers the forward region of the interaction point, from 30 mrad to 300 (250) mrad in the bending (non-bending) plane. The choice of such a limited acceptance is motivated by the fact that most of the $\approx 500\,\mu$b correlated $b\bar{b}$ pairs are produced in this region. LHCb experiment will take data at a luminosity of $2 \times 10^{32} \mathrm{cm}^{-2}\mathrm{s}^{-1}$, where bunch crossing are dominated by single pp interactions. Good particle identification, excellent tracking and vertexing are needed for B physics measurements. Expected resolution on track momentum is about $\delta\mathrm{p}/\mathrm{p} = 0.35\%$ around 10 GeV/c to $\delta\mathrm{p}/\mathrm{p} = 0.55\%$ around 140 GeV/c. Impact parameter resolution is expected to be $\sigma_{\mathrm{IP}} = 14\mu\mathrm{m} + 35\mu\mathrm{m}/\mathrm{p_T}$.

5.1 Reconstruction and identification of b-jets

As a textbook case, we study in the following the case of a Higgs boson decaying into $b\bar{b}$ pairs, produced in association with a vector boson decaying leptonically. The Higgs mass is chosen
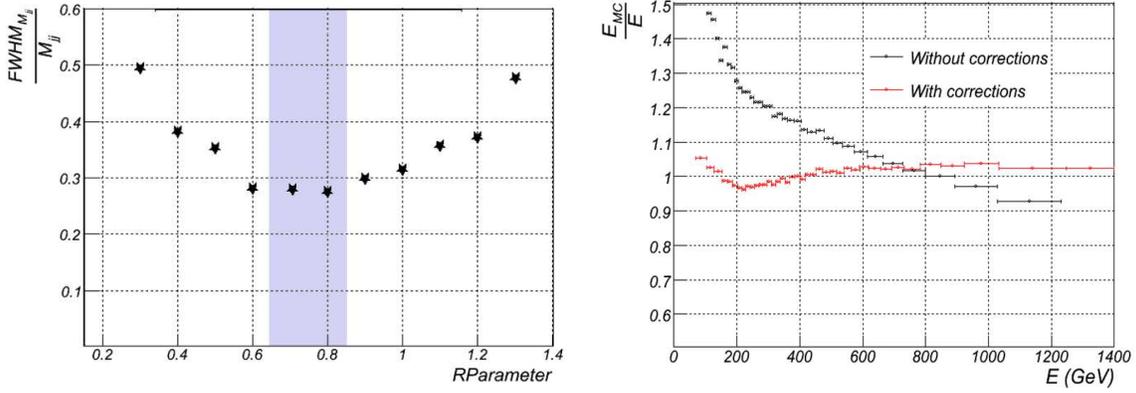
Fig. 15: Full width at half maximum (FWHM) over peak value of the di-jet mass distribution considering all particles from the generator, except neutrinos (left). True jet energy over reconstructed jet energy as a function of the reconstructed jet energy (right).

to be $120\,\mathrm{GeV/c^2}$ and the lepton, with a transverse momentum p_T higher than $10\,\mathrm{GeV/c}$, is required to be in the LHCb acceptance.

Several contributions might affect the di-jet mass resolution. In order to choose the best working point for the jet algorithm, a generator level study of the di-jet mass reconstruction is performed. The width of the di-jet mass distribution is shown in Fig. 15. At small R^2, gluon radiation and hadronisation induce a low-mass tail. This effect tends to increase the width of the distribution. For large values of R, the area of the jets is larger and the contribution of underlying event particles increases. This effect induces high-mass tails, increasing again the width of the distribution. Taking into account both effects, the optimal R value is between 0.7 and 0.9. The limited acceptance of the detector is the major restriction for jet studies at LHCb. Some jet particles arriving close to the detector border are not reconstructed. Therefore these jets need to be rejected.

Using the LHCb reconstruction framework, charged and neutral particles are separated by matching tracks and electromagnetic calorimeter clusters. In the full simulation, track measurements are used for charged particles while the energy of electromagnetic clusters is used for neutral particles. Tracks and electromagnetic clusters are calibrated objects. In a simple approach, clusters in the hadronic calorimeter are not used. The charged and neutral particles are used as input to the jet algorithm. Two ways of reconstructing and identifying b-jets are under study. The first way consists in finding tracks coming from B-decays, and use them as seed for cone-type jet finding. The second way uses a sequential recombination algorithm, k_t-jet [1], to reconstruct the jets. the k_t-jet algorithm with an R parameter of 0.75 gives on average 15 (proto-)jets for our textbook case events. Only a quarter of them has a transverse momentum higher than 5 GeV/c. The performance of b-jets reconstruction is presented below.

A primary cut is applied to remove (uninteresting) jets with $p_T < 5\,\mathrm{GeV/c}$, less than 4 constituents and 2% of charged energy. The content of jets in particles from B-decays is

---

[2]R is a generic parameter of the jet algorithm, representing in first approximation the (η, ϕ) radius.

quantified by two weights separating b- from c- and light-jets. To construct these weights, tri-dimensional probability distribution functions have been extracted from Monte Carlo (MC) for particles that are respectively b-, light- and c-jet constituents. b-jet identification is then based on combination of these weights with several variables such as the number of constituents, the energy contained in a cone of $R < 0.4$ around the jet axis, the charged energy percentage, etc. Taking only into account the jets that pass the primary cut, b-jet selection efficiency is about 81%, for a rejection of about 95% of light-jets and 91% of c-jets in $t\bar{t}$ events. It is interesting to notice that the b-jets only partially inside the acceptance are rejected at more than 90% by the selection. At this level, no explicit reconstruction of vertices, nor semi-leptonic decay identification has been performed. This leaves room for future improvements.

Reconstruction of b-jets is efficient in the range of pseudorapidity $2 < \eta < 4.2$. For jets well contained in the acceptance, one can determine corrections to the jet energy depending on its p_T and pseudorapidity. The corrections are determined from a $t\bar{t} \to b\bar{b} + \ell$ MC sample and applied to the b-jets of the MC Higgs sample. After correction, the energy response presents a non-linearity below 5%, see Fig. 15. In a Gaussian approximation, the di-jet mass resolution is improved from $\sigma/\text{mean} = 24\%$ to 20%. The mass pick is still offset. The contribution due to loss of neutral hadrons will be added.

The contribution, after b-jets selection, of the remaining "partially inside the acceptance" b-jets, is shown Fig. 16. There is a 10% increase in the resolution of the corrected di-jet mass distribution due to the pollution of those jets.

5.2 Interesting processes for LHCb

A measurement of $H(W, Z) \to b\bar{b} + \ell$ would be very interesting. But the level of background, especially $t\bar{t} \to b\bar{b} + \ell$ is large and its suppression is a real challenge for LHCb. Because of the limited acceptance, global event variables (*e.g.* missing E_T, sphericity, etc.) are inaccessible. But this analysis also gives the opportunity to develop tools for b-jets studies that might be of interest for other studies involving b-jets in the forward region.

Many new physics models give rise to particles with measurable lifetime decaying into b quarks. In the following we will concentrate on feasibility of two such models.

Hidden valley is a class of phenomenological models that extends the Standard Model (SM) gauge group *GSM* with a non-abelian group *Gv*. High-dimension operators at the TeV scale allow interactions between SM and new particles. Some neutral v-hadrons, π_v, can decay into the gauge-invariant combinations of SM-particles with observable lifetimes. An interesting process is described by M.J.Strassler and K.M.Zurek [33]. The SM Higgs, thanks to the coupling to a new scalar field, decays into two π_v, which decay into $b\bar{b}$ pairs. For $m_{\pi_v} = 45 \text{ GeV}/c^2$, the probability of correct reconstruction of such a b-jet in LHCb is about 30% up to 50 cm flight path of the π_v. The di-jet mass of b-jets from π_v decay is shown Fig. 17.

Another interesting model was developed by L. Carpenter, D. Kaplan and E-J. Rhee [34]. They have shown that the MSSM with R-parity violation, baryon number violation and non-unified gaugino masses has a non-excluded parameter space in which a light boson decays mainly into two neutralinos. The neutralino decay length varies as the inverse square of the baryon num-

ber violation coupling constant λ''. The final state of such events contains six quarks, among which the probability to find b and c is large. The two vertices from the $\tilde{\chi}^0 \rightarrow (b, c)$ sequence are reconstructed and assembled to get the $\tilde{\chi}^0$ vertex mass. A study at generator level with vertex smearing has been performed. For $m_{\tilde{\chi}^0} = 50\,\mathrm{Gev/c^2}$, $m_{h^0} = 115\,\mathrm{Gev/c^2}$ and $\lambda'' = 10^{-4}$, about 16600 events of signal are expected. After selection of the 4 vertices per event, one gets $\approx$6000 events, and most of the background is rejected. The h^0 mass distribution after selection, with the remaining events of $b\bar{b}$, $t\bar{t}$ and $Z^0 W^\pm \rightarrow b\bar{b}$, is shown on Fig. 17. Studies of vertex reconstruction and background rejection with full simulation are ongoing. Details can be found in [35].

It has been shown that LHCb can reconstruct b-jets in the forward region ($2 < \eta < 4$), and reconstruct the di-jet mass with a resolution of about 20%. Selection of b-jets benefits a lot from the quality of the LHCb apparatus. Besides important B physics measurements, LHCb has the potential to observe new physics processes in the high rapidity region looking at b-jets and highly displaced vertices.

I would like to thank the LHCb Collaboration and the jets working group for stimulating discussions on the subject and for their help in the preparation of this talk.

References

[1] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, Nucl. Phys. **B406**, 187 (1993).

[2] S. D. Ellis and D. E. Soper, Phys. Rev. **D48**, 3160 (1993), `arXiv:hep-ph/9305266`.

[3] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, JHEP **08**, 001 (1997), `hep-ph/9707323`.

[4] M. Wobisch and T. Wengler (1998), `hep-ph/9907280`.

[5] G. C. Blazey *et al.* (2000), `arXiv:hep-ex/0005012`.

[6] G. P. Salam and G. Soyez, JHEP **05**, 086 (2007), `arXiv:0704.0292 [hep-ph]`.

[7] G. P. Salam and G. Soyez, SISCone, `http://projects.hepforge.org/siscone`.

[8] M. Cacciari and G. P. Salam, Phys. Lett. **B641**, 57 (2006), `arXiv:hep-ph/0512210`.

[9] G. P. S. M. Cacciari and G. Soyez, `www.lpthe.jussieu.fr/~salam/fastjet`.

[10] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **04**, 063 (2008), `arXiv:0802.1189 [hep-ph]`.

[11] T. Sjostrand, S. Mrenna, and P. Skands, JHEP **05**, 026 (2006), `arXiv:hep-ph/0603175`.

[12] C. Buttar *et al.* (2008), `arXiv:0803.0678 [hep-ph]`.

[13] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock, and B. R. Webber, Phys. Lett. **B269**, 432 (1991).

[14] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, Phys. Rev. Lett. **100**, 242001 (2008), `arXiv:0802.2470 [hep-ph]`.

[15] M. Cacciari, J. Rojo, G. P. Salam, and G. Soyez, JHEP **12**, 032 (2008), `arXiv:0810.1304 [hep-ph]`.

[16] M. Dasgupta, L. Magnea, and G. P. Salam, JHEP **02**, 055 (2008), `arXiv:0712.3014 [hep-ph]`.

[17] A. Bhatti *et al.* (2008), `arXiv:0807.4961 [hep-ex]`.

[18] M. Cacciari and G. P. Salam, Phys. Lett. **B659**, 119 (2008), `arXiv:0707.1378 [hep-ph]`.

[19] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **04**, 005 (2008), `arXiv:0802.1188 [hep-ph]`.

[20] T. C. Collaboration, CMS PAS JME-07-003 (2008).

[21] ATLAS Collaboration, S. Bentvelsen *et al.*, JINST **3**, S08003 (2008).

[22] A. Collaboration, *Expected Performance of the ATLAS Experiment, Detector, Trigger and Physics – Jets and Missing Et Chapter, Jet Reconstruction Performance*, No. CERN-OPEN-2008-020. CERN, 2008, to appear.

[23] W. e. a. Lampl, *Calorimeter Clustering Algorithms: Description and Performance*, Technical Report ATL-LARG-PUB-2008-002, CERN, Geneva, Apr 2008.

[24] A. Collaboration, *Expected Performance of the ATLAS Experiment, Detector, Trigger and Physics – Jets and Missing Et Chapter, Detector Level Jet Corrections*, No. CERN-OPEN-2008-020. CERN, 2008, to appear.

[25] A. Collaboration, *Expected Performance of the ATLAS Experiment, Detector, Trigger and Physics – Jets and Missing Et Chapter, E/p Performance for Charged Hadrons*, No. CERN-OPEN-2008-020. CERN, 2008, to appear.

[26] A. Collaboration, *Expected Performance of the ATLAS Experiment, Detector, Trigger and Physics – Jets and Missing Et Chapter, In-situ Calibration Strategies*, No. CERN-OPEN-2008-020. CERN, 2008, to appear.

[27] A. Collaboration, *Expected Performance of the ATLAS Experiment, Detector, Trigger and Physics – Top-quark Physics at ATLAS CSC Chapter*, No. CERN-OPEN-2008-020. CERN, 2008, to appear.

[28] A. P. Cheplakov and S. Thompson, *MidPoint Algorithm for Jets Reconstruction in ATLAS Experiment*, Technical Report ATL-PHYS-PUB-2007-007, CERN, Geneva, Dec 2006.

[29] D. Y. Grigoriev, E. Jankowski, and F. V. Tkachov, Comput. Phys. Commun. **155**, 42 (2003), `arXiv:hep-ph/0301226`.

[30] J. M. Butterworth, A. Davison, E. Ozcan, and P. Sherwood, *YSplitter: An Athena tool for studying jet substructure*, Technical Report ATL-PHYS-INT-2007-015. ATL-COM-PHYS-2007-077, CERN, Geneva, Oct 2007.

[31] A. Collaboration, *Expected Performance of the ATLAS Experiment, Detector, Trigger and Physics*, No. CERN-OPEN-2008-020. CERN, 2008, to appear.

[32] LHCb Collaboration, S. Amato *et al.* CERN-LHCC-98-04.

[33] M. J. Strassler and K. M. Zurek, Phys. Lett. **B651**, 374 (2007), `arXiv:hep-ph/0604261`; M. J. Strassler and K. M. Zurek, Phys. Lett. **B661**, 263 (2008), `arXiv:hep-ph/0605193`.

[34] D. E. Kaplan and K. Rehermann, JHEP **10**, 056 (2007), `arXiv:0705.3426 [hep-ph]`; L. M. Carpenter, D. E. Kaplan, and E.-J. Rhee, Phys. Rev. Lett. **99**, 211801 (2007), `arXiv:hep-ph/0607204`.
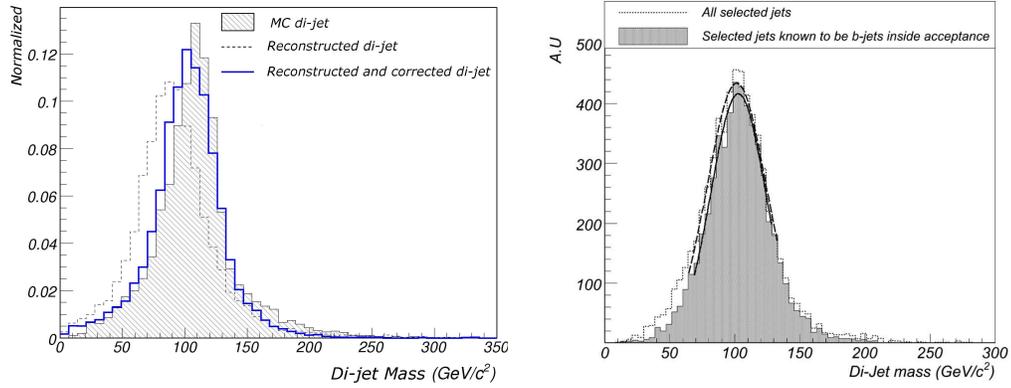
[35] N. Gueissaz. CERN-THESIS-2007-038.

Fig. 16: Full simulation di-jet mass with and without correction, (left). Corrected di-jet mass with "inside the acceptance" b-jets only (filled grey), and all selected b-jets (dashed line), (right).
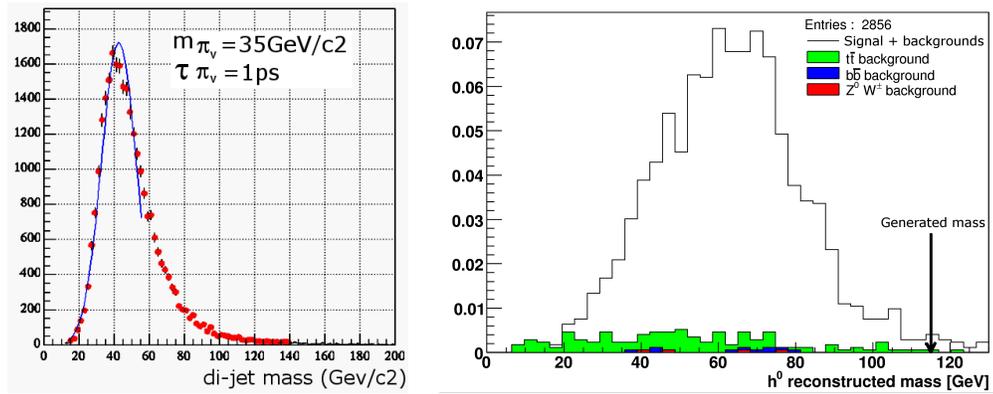

Fig. 17: Reconstructed π_v mass from Higgs decays with $m_H = 120 \, \text{GeV}/c^2$ in the hidden valley model (left). Higgs mass reconstructed from the two $\tilde{\chi}^0$ vertices in MSSM with an R-parity violation model (right).