Review on recent developements in jet finding

Juan Rojo
LPTHE, UPMC – Paris 6 and Paris-Diderot – Paris 7, CNRS UMR 7589, Paris (France)
INFN, Sezione di Milano, Via Celoria 16, I - 20133, Milano (Italy)

DOI: http://dx.doi.org/10.3204/DESY-PROC-2009-01/49

Abstract
We review recent developements related to jet clustering algorithms and jet finding. These include fast implementations of sequential recombination algorithms, new IRC safe algorithms, quantitative determination of jet areas and quality measures for jet finding, among many others. We also briefly discuss the status of jet finding in heavy ion collisions, where full QCD jets have been measured for the first time at RHIC.

Recent developements in jet algorithms With the upcoming start-up of the LHC, jet finding techniques have received considerable attention. In this brief review, we outline some of the most important developements in jet algorithms and related subjects in the recent years. Much more detailed reviews can be found in [1, 2].

An important developement has been the fast implementation of the k_T [3] and Cambridge/Aachen [4, 5] jet algorithms. Prior to 2005, existing implementations scaled as N^3, with N the number of particles to be clustered, thus making it unpractical for high multiplicity collisions like pp at the LHC and even more in Heavy Ions Collisions (HIC). Thanks to computational geometry methods, the performance of these algorithms was made to scale as $N \ln N$ [6]. These fast implementations are available through the `FastJet` package [7], together with area-based subtraction methods and plugins to external jet finders (see below).

Another important achievement has been the formulation of a practical (scaling as $N^2 \ln N$) infrared and collinear (IRC) safe cone algorithm, SISCone [8]. Unlike all other commonly used cone algorithms, SISCone is IRC safe to all orders in perturbation theory by construction. This property allows one to compare any perturbative computation with experimental data, which for IRC unsafe algorithms is impossible beyond some fixed order, indicated in Fig. 2. As discussed in [8], the phenomenological implications of SISCone when compared with the (IRC unsafe) commonly used MidPoint cone algorithm range from few percent differences in the inclusive jet spectrum, somewhat larger in the presence of realistic Underlying Event (UE), up to 50% differences for more exclusive observables, like the tails of jet-mass spectra in multi-jet events.

There has been historically some confusion about the concept of the *size* of a jet, specially since the naive jet area is ambiguous beyond LO. The situation was recently clarified by the introduction of quantitative definitions of jet areas based on the *catchment* properties of hard jets with respect to very soft particles, called *ghosts* in [9]. Examples of jet areas defined with such a technique are shown in Fig. 1. On top of their theoretical importance, jet areas have important applications related to the subtraction of soft backgrounds coming from the UE or from Pile-Up (PU), both in pp and in AA collisions, as discussed in [10].

Another recently developed IRC safe jet algorithm is the anti-k_t algorithm [11]. This algorithm is related to k_T and Cam/Aa by its distance measure, $d_{ij} \equiv \min(k_{ti}^{2p}, k_{tj}^{2p})\Delta R_{ij}^2/R^2$, with $p = -1$ ($p = 1$ corresponds to k_T and $p = 0$ to Cam/Aa). The anti-k_T algorithm has the property of being soft-resilient, that is, due to its distance soft particles are always clustered with hard particles first. This property leads to rather regular jet areas, which become perfectly circular in the limit in which all hard particles are separated in the (y, ϕ) plane by at least a distance R, as can be seen in Fig. 1. Another important advantage of the anti-k_t algorithm is that it has a very small back-reaction [9], that is, the presence of a soft background has reduced effects on which hard particles are clustered into a given jet.

The recent progress in jet algorithms can be summarized in Fig. 2. Each IRC unsafe cone jet algorithm can now be replaced by the corresponding IRC safe one, with a similar physics performance, shown in the last column of Fig. 2. SISCone is the natural IRC safe replacement for MidPoint-type iterative cone algorithms with split-merge (IC-SM), while anti-k_T is so for iterative cone algorithms of the progressive removal (IC-PR) type [1].

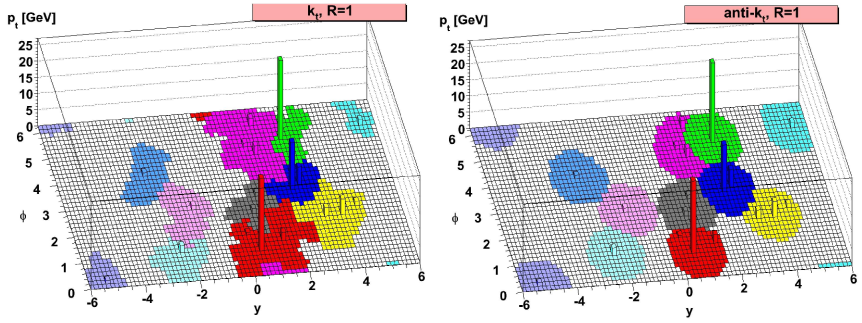

Fig. 1: Jet areas for the k_t (left) and anti-k_t (right) algorithms for $R = 1$.

This brief review is unable to cover many other interesting developements related to jets and jet finding in the recent years. Some of those not discussed here include the use of jet substructure as a useful technique to improve signal significance in various channels at the LHC (see for example [12–14]), analytical studies of the interplay between perturbative and non-perturbative effects in jet finding [15], the infrared safe definition of jet flavour and its application to precision predictions for $b-$jets at hadron colliders [16, 17] or the impact of jet measurements, both at the Tevatron and at HERA, in global analysis of PDFs [18, 19].

Performance of jet algorithms at LHC A recurring question in jet studies is "what is the best jet definition for a given specific analysis"? Most existing techniques either use as a reference unphysical Monte Carlo partons (an ambiguous concept beyond LO) and/or assume some shape for the measured kinematical distributions. To overcome these disadvantages, a new strategy to quantify the performance of jet definitions in kinematic reconstruction tasks has been recently introduced [20], which is designed to make use exclusively of physical observables.

In Ref. [20] two quality measures respecting the above requirements are proposed, and applied to the kinematic reconstruction of invariant mass distributions in dijet events for a wide

Algorithm	Type	IRC status	Evolution
exclusive k_t	$\text{SR}_{p=1}$	OK	$N^3 \to N \ln N$
inclusive k_t	$\text{SR}_{p=1}$	OK	$N^3 \to N \ln N$
Cambridge/Aachen	$\text{SR}_{p=0}$	OK	$N^3 \to N \ln N$
Run II Seedless cone	SC-SM	OK	$\to$ SISCone
CDF JetClu	IC_r-SM	IR_{2+1}	[$\to$ SISCone]
CDF MidPoint cone	IC_{mp}-SM	IR_{3+1}	$\to$ SISCone
CDF MidPoint searchcone	$\text{IC}_{se,mp}$-SM	IR_{2+1}	[$\to$ SISCone]
D0 Run II cone	IC_{mp}-SM	IR_{3+1}	$\to$ SISCone [with p_t cut?]
ATLAS Cone	IC-SM	IR_{2+1}	$\to$ SISCone
PxCone	IC_{mp}-SD	IR_{3+1}	[little used]
CMS Iterative Cone	IC-PR	Coll_{3+1}	$\to$ anti-k_t
PyCell/CellJet (from Pythia)	FC-PR	Coll_{3+1}	$\to$ anti-k_t
GetJet (from ISAJET)	FC-PR	Coll_{3+1}	$\to$ anti-k_t

SR = seq.rec.; IC = it.cone; FC = fixed cone;
SM = split–merge; SD = split–drop; PR = progressive removal

Fig. 2: Summary of the progress in jet algorithms since 2005.

range of energies. These quality measures can in turn be mapped into an effective luminosity ratio, defined as

$$\rho_{\mathcal{L}}(\text{JD}_2/\text{JD}_1) \equiv \frac{\mathcal{L}(\text{needed with JD}_2)}{\mathcal{L}(\text{needed with JD}_1)} = \left[\frac{\Sigma\,(\text{JD}_1)}{\Sigma\,(\text{JD}_2)}\right]^2 . \tag{1}$$

Given a certain signal significance Σ with jet definition JD_2, $\rho_{\mathcal{L}}(\text{JD}_2/\text{JD}_1)$ indicates the factor more luminosity needed to obtain the same significance as with jet definition JD_1.

The results of [20] over a large range of jet definitions,[1] summarized in Fig. 3, indicate that for gluon jets, and in general for TeV scales, there are significant benefits from using larger radii that those commonly used, up to $R \gtrsim 1$. In general, SISCone and C/A-filt (Cam/Aa supplemented with a filtering procedure [12]) show the best performance. These conclusions are robust in the presence of high-luminosity PU, when subtracted with the jet area technique [10].

Jet finding in AA collisions at RHIC and LHC While QCD jets are ubiquitous in pp collisions, until this year no real jet reconstruction had been obtained in the much more challenging environment of HIC. Indeed, usually in HIC one refers to the leading particle of the event as a *jet*. However, reconstructing full QCD jets provides a much more precise window to the properties of the hot and dense medium created in the collision than just leading particles.

The difficulty in reconstructing jets in HIC stems from the huge backgrounds, which need to be subtracted in order to compare with baseline results. There are various techniques to subtract such large backgrounds. In [10] it was shown how the `FastJet` jet area method could efficiently subtract such backgrounds in HIC at the LHC with a good accuracy (see Fig. 4).

A major breakthrough in jet finding was the recent first measurement of QCD jets in HIC by the STAR collaboration at RHIC [22]. In Fig. 4 we show their measurement with the k_T

[1] There results can also be accessed through an interactive web tool [21] which allows the user to compare the jet finding quality for a wide range of parameters (jet algorithm, R, value of PU, ...).

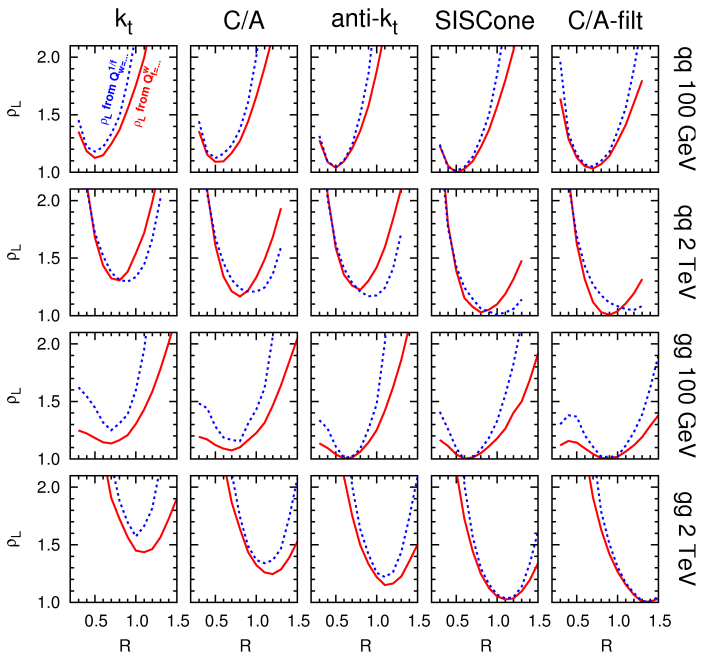

Fig. 3:
The effective luminosity ratio, Eq. 1, for quark and gluon jets at 100 GeV and 2 TeV [20].

algorithm. These results should have important consequences for understanding the medium properties in HIC.

It would be important, after these initial measurements, to improve the control on the accuracy of the subtraction procedure, as well as to understand the differences between the performances of different jet algorithms. Ongoing studies [23] suggest that one of the important sources of systematic error in the HIC jet reconstruction is back-reaction [9], therefore anti-k_t is potentially interesting in this situation due to its small back-reaction [11]. Ref. [23] also investigates how the use of local ranges for the determination of the background level ρ might help reducing the effects of point-to-point background fluctuations. However, more work is still required in order to determine the optimal settings for jet finding in HIC.

Outlook Jet finding has seen a large number of important developements in the recent years, However, there is still room for more progress, which should be driven by the actual requirements of LHC data analysis. Jet finding will also be essential to exploit the heavy-ion program at the LHC as proved by the latest RHIC jet measurements.

Acknowledgments This work has been supported by the grant ANR-05-JCJC-0046-01 (France). The author wants to acknowledge M. Cacciari, G. Salam and G. Soyez for help and material while preparing this review.

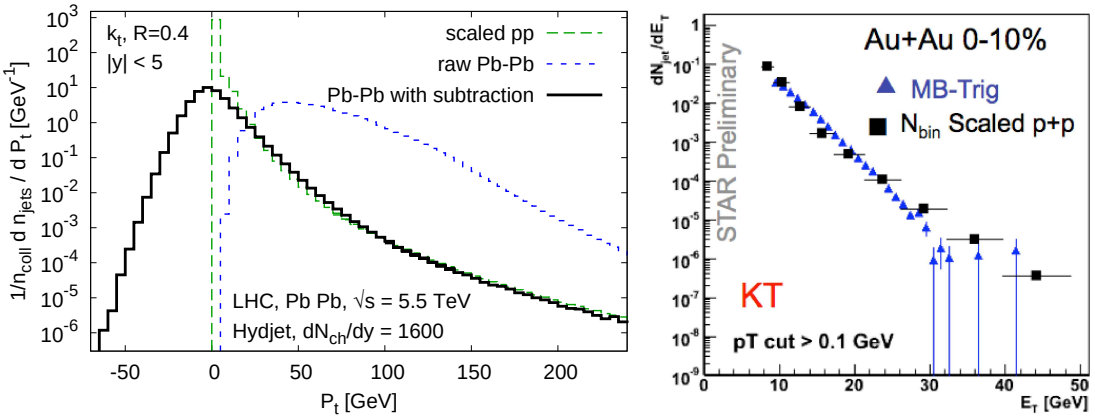

Fig. 4:
Left: the simulated inclusive jet spectrum at the LHC with the k_T algorithm, including subtraction, from [10]. Right: the inclusive jet spectrum measured with k_T by STAR at RHIC, from [22].

References

[1] C. Buttar *et al.* (2008). 0803.0678 [hep-ph].

[2] S. D. Ellis, J. Huston, K. Hatakeyama, P. Loch, and M. Tonnesmann, Prog. Part. Nucl. Phys. **60**, 484 (2008). 0712.2447.

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

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

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

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

[7] M. Cacciari, G. P. Salam, and G. Soyez (2005-2008). http://www.lpthe.jussieu.fr/∼salam/fastjet.

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

[9] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **04**, 005 (2008). 0802.1188.

[10] M. Cacciari and G. P. Salam, Phys. Lett. **B659**, 119 (2008). 0707.1378.

[11] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **04**, 063 (2008). 0802.1189.

[12] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, Phys. Rev. Lett. **100**, 242001 (2008). 0802.2470.

[13] D. E. Kaplan, K. Rehermann, M. D. Schwartz, and B. Tweedie, Phys. Rev. Lett. **101**, 142001 (2008). 0806.0848.

[14] J. Thaler and L.-T. Wang, JHEP **07**, 092 (2008). 0806.0023.

[15] M. Dasgupta, L. Magnea, and G. P. Salam, JHEP **02**, 055 (2008). 0712.3014.

[16] A. Banfi, G. P. Salam, and G. Zanderighi, Eur. Phys. J. **C47**, 113 (2006). hep-ph/0601139.

[17] A. Banfi, G. P. Salam, and G. Zanderighi, JHEP **07**, 026 (2007). 0704.2999.

[18] NNPDF Collaboration, R. D. Ball *et al.* (2008). 0808.1231.

[19] P. M. Nadolsky *et al.*, Phys. Rev. **D78**, 013004 (2008). 0802.0007.

[20] M. Cacciari, J. Rojo, G. P. Salam, and G. Soyez (2008). 0810.1304.

[21] M. Cacciari, J. Rojo, G. P. Salam, and G. Soyez (2008), http://quality.fastjet.fr/.

[22] STAR Collaboration, S. Salur (2008). 0809.1609.

[23] M. Cacciari, J. Rojo, G. P. Salam, and G. Soyez , in preparation.