RECENT PROGRESS IN DEFINING AND UNDERSTANDING JETS*

Gavin P. Salam

LPTHE, CNRS UMR 7589
Université Pierre et Marie Curie (Paris VI)
Université Denis Diderot (Paris VII)
75252 Paris Cedex 05, France

(Received October 16, 2008)

This talk reviews some key developments that have taken place in hadron-collider jet finding over the past couple of years, including: technical advances such as the complete formulation of an infrared safe seedless cone algorithm and fast computational approaches to sequential recombination jet finders like the k_t algorithm, together with universal methods for subtracting pileup; progress in understanding the sensitivity of jet algorithms to the underlying event and hadronisation; and work that exploits our knowledge of QCD divergences to better define and predict heavy-flavour jet cross sections.

PACS numbers: 13.87.–a

1. Introduction

Jet algorithms provide a way of projecting away the multiparticle dynamics of an event so as to leave a simple quasi-partonic picture of the underlying hard scattering. This projection is however fundamentally ambiguous, reflecting the divergent and quantum mechanical nature of QCD. Consequently, jet physics is a rich subject.

Key developments in the history of jet finding have often been spurred by advances in experimental sophistication, and in this vein, the upcoming startup of the LHC provides a motivation for reexamining the technology at our disposal.

To appreciate what changes at LHC, consider the physics scales and processes at play: in addition to having the electroweak (~ 100 GeV) and hadronisation (0.5 GeV) scales familiar at LEP and HERA, and an underlying event (~ 10 GeV) 2–4 times larger than the Tevatron's, the LHC will

* Presented at the XXXVII International Symposium on Multiparticle Dynamics, Berkeley, USA, August 4–9, 2007.

routinely probe multi-jet events of unprecedented complexity (think $t\bar{t}H \to 8$ jets), it will suffer from huge pileup (~ 100 GeV of $p_{\rm t}$ per unit rapidity), and it may well discover new-particle cascades that mix the TeV scale and electroweak scales. That is vastly more to disentangle than ever before. To add to that, a key technical issue is posed by the number of particles: up to ~ 4000 per event, two orders of magnitude larger than at LEP (~ 50), and an order larger than at Tevatron.

A programme of work to bring jet-finding up-to-date for the LHC age has begun over the past couple of years and involves three main phases: (1) develop a core set of theoretically solid and experimentally practical jet algorithms (*i.e.* the 1990 Snowmass accord [1]); (2) quantify, where possible analytically, how jet algorithms respond to various non-perturbative and perturbative QCD effects; (3) use the resulting understanding to guide development of more sophisticated tools. As described below, phase 1 is nearing completion, and progress is being made on the remaining parts.

2. Core tools

The 1990 Snowmass accord [1] for the Tevatron advocated the use of jet algorithms that were simple to use, both theoretically and experimentally, well-defined and finite at all orders of perturbation theory, and relatively insensitive to hadronisation. However, this accord has never fully been satisfied in Tevatron jet finding.

Cone algorithms provide a top-down form a jet identification, and are mostly based on the idea of a *stable cone*, one whose direction coincides with that of the summed momenta of the contained particles. They are widespread at *pp* colliders, motivated on the grounds that soft and collinear radiation leaves stable cones unchanged, and a feature often quoted as being one of their main experimental advantages is their simple conical shape.

Cone algorithms have long been plagued by infrared and collinear (IRC) safety issues. Old iterative cones with split-merge procedures are unreliable from order $\alpha_{\rm s}^3$ (or $\alpha_{\rm EW}\alpha_{\rm s}^2$) onwards, Tevatron Run II "midpoint" types cones from order $\alpha_{\rm s}^4$ (or $\alpha_{\rm EW}\alpha_{\rm s}^3$). Unreliable at order $\alpha_{\rm s}^n$ means here that they diverge when calculated at $\alpha_{\rm s}^{n+1}$ — regulating that divergence around $\Lambda_{\rm QCD}$ introduces a term $\sim \alpha_{\rm s}^{n+1} \ln p_{\rm t}/\Lambda_{\rm QCD}$. This is the same size as the $\alpha_{\rm s}^n$ term (recall $\alpha_{\rm s} \sim 1/\ln(p_{\rm t}/\Lambda_{\rm QCD})$), *i.e.* any effort[1] that went into the $\alpha_{\rm s}^n$ precision is swamped by the near-divergent uncalculated higher orders.

The practical relevance of the IRC safety issue has been repeatedly questioned, it being noted for example that in the most widely studied of jet-observables, the inclusive-jet spectrum, the "real" effect of IR unsafety is seen to be 1%. This figure, however, holds only for this observable: leading

[1] Between 50 and 100 people working over ten years, *i.e.* $\sim$ \$50 million.

order for the jet-spectrum is α_s^2, the midpoint cone's unreliability starts at α_s^4, and for $\alpha_s \simeq 0.1$ the ratio is 1%. At LHC, many interesting processes *start* at α_s^4 or higher, and then even leading order can be unreliable, with up to 50% effects, depending on the cuts.

The design of an IRC safe cone algorithm starts with the observation that you should find *all* stable cones [2]. Ref. [3] showed how, for a handful of particles (in $N2^N$ time, *i.e.* 10^{17} years for $N = 100$). Recently, Ref. [4] reduced that to a more manageable $N^2 \ln N$. The trick was to recast it as a computational geometry problem, *i.e.* the identification of all distinct circular enclosures for points in 2D, and to find a (previously unknown) solution to that. Together with a few other minor fixes, this has led to the first ever IRC safe jet algorithm, SISCone (*cf.* left plot of Fig. 1).

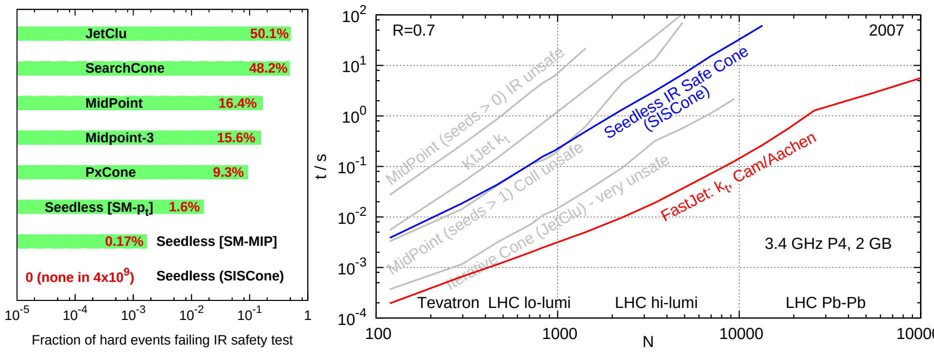

Fig. 1. Left: IR safety failure rate for a range of jet algorithms in artificial events with between 2 and 10 hard particles (for details, see [4]). Right: speeds of various algorithms as a function of the particle multiplicity N.

Sequential recombination algorithms (SRAs), such as k_t [5], take a bottom-up approach to creating jets, successively merging the closest pair of objects in an event until all are sufficiently well separated. They work because of relations between the distance measures that are used and the divergences of QCD. Their attractions include their conceptual simplicity, as well as the hierarchical structure they ascribe to an event, and they were ubiquitous at LEP and HERA.

There had been two major issues for SRAs in *pp* collisions: they used to be slow ($\sim N^3$ time to cluster N particles, *i.e.* 1 minute for $N = 4000$) and the shape of the resulting jets was unknown and irregular, which complicated pileup subtraction. Recently the speed issue was solved [6] by observing a connection with computational geometry problems: *e.g.* the k_t algorithm factorises into a priority queue and the problem of constructing a nearest-neighbour-graph in 2D and maintaining it under point changes (solved in [7]). Asymptotically, run times are now $N \ln N$, and in practice ~ 20 ms for $N = 4000$. That's better even than a fast (but very IR unsafe) iterative cone algorithm such as CDF's JetClu (*cf.* right-plot of Fig. 1).

The problem of the unknown shape of SRA jets has also been solved, by the simple expedient of adding very many infinitely soft "ghost" particles [8]. These serve to fill in all empty space in the event and so give a well-defined boundary and total area to each jet. Subtracting a correction proportional to that area works rather well for removing pileup [9] and can even be applied to the extremely noisy environment of LHC Pb Pb collisions (Fig. 2). This progress, together with the recent successful measurement of the inclusive jet spectrum by CDF with the k_t algorithm [10], means that all objections raised in the past about SRAs are now essentially resolved.

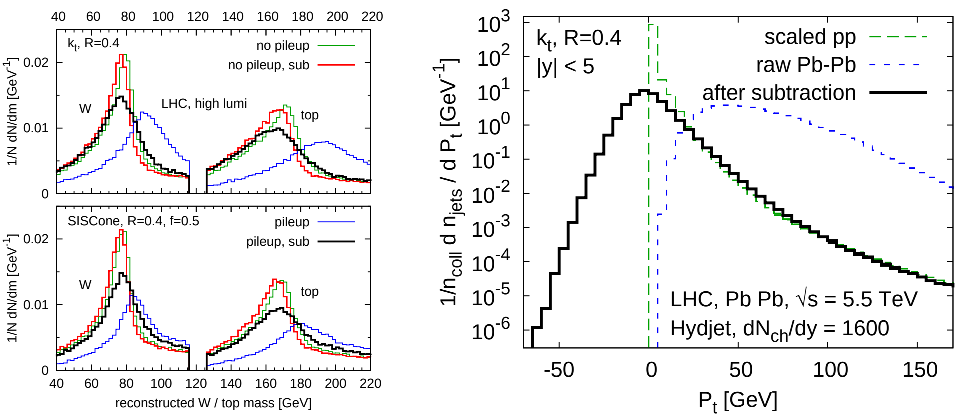

Fig. 2. Left: hadronic W and top mass peaks with pileup subtraction in high-luminosity semi-leptonic $t\bar{t}$ events at LHC. Right: inclusive jet spectrum in LHC PbPb events before and after background subtraction. Adapted from [9], simulations used Pythia [11] and Hydjet [12].

3. Understanding and improving jet algorithms

Once you have a set of safe, fast algorithms (all conveniently packaged in `FastJet` [6]), you can start trying to understand their physics behaviour. A simple question, for example, is that of how **hadronisation and the underlying event** (UE) modify a jet's transverse momentum. The situation is summarised in Table I [13], whose results are essentially common to all jet algorithms with a radius parameter R. The distinct R-dependence for each effect may provide a way of disentangling them experimentally. It also implies an optimal R (minimising the sum of squares of effects) that varies significantly with the jet initiator's colour and p_t, as illustrated in Fig. 3, based on the results in Table I: at large p_t perturbative radiation dominates over other contributions, so one prefers $R \sim 1$, whereas at low p_t the UE has a significant relative impact and it is advantageous to lower R to limit this, especially at higher energy colliders (since the UE grows with s) and for quark jets (for which perturbative radiation is weaker). Note that these results for the optimal R are mainly to be taken as indicative of general trends: a definitive estimate would go beyond the small-R approximation and take into account the dispersion of each effect rather than its mean value.

TABLE I

Summary of the main physical effects that contribute to the average difference $\langle \delta p_t \rangle$ between the transverse momentum of a jet and its parent parton (for small R). $\Lambda_h \simeq 0.35 - 0.4$ GeV based on e^+e^- event-shape studies [14], $\Lambda_{UE} s^{\omega} \simeq 4\,\text{GeV} \times \left(s/(2\,\text{TeV})^2\right)^{0.25}$ [13].

	Jet $\langle \delta p_t \rangle$ given by product of dependence on			
	Scale	Colour factor	R	$\sqrt{s}$
Perturbative radiation	$\sim \frac{\alpha_s(p_t)}{\pi} p_t$	C_i	$\ln R + \mathcal{O}(1)$	—
Hadronisation	Λ_h	C_i	$-1/R + \mathcal{O}(R)$	—
Underlying event	Λ_{UE}	—	$R^2/2 + \mathcal{O}(R^4)$	s^{ω}

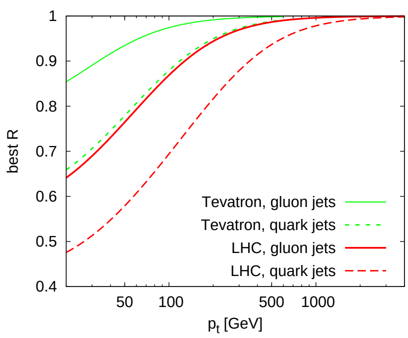

Fig. 3. Simple estimate for optimal R as a function of jet p_t, collider and initiating parton.

While to a first approximation the effects shown in Table I are independent of the specific jet algorithm, more refined studies, for jet areas [8], do highlight differences between algorithms, but not always as one would expect. For example, with heavy pileup, the k_t algorithm, often labelled a vacuum cleaner, actually has an average area quite close to πR^2 (essentially because pileup is not vacuum); cone algorithms are widely assumed to have an area πR^2, but modern versions with split-merge steps (*e.g.* SISCone) actually turn out not to be quite conical, having an area $\sim \pi R^2/2$. This *small* area is part of the reason why they work well in noisy environments[2]. This has important implications for strategies that assume an area of πR^2 in correcting for pileup with cone-type algorithms.

Perhaps the most striking example to date where a better understanding of clustering dynamics can lead to improved algorithms concerns **jet flavour**. This concept is often taken for granted (over 350 articles' titles contain the words "quark jet" or "gluon jet"), and it would seem that if one simply sums the flavours of all partons in a jet one might obtain a well-defined result for the jet-flavour. This turns out not to be the case, even in algorithms for which the jet momenta are IRC safe, because the flavour is subject to contamination by large-angle $g \to q\bar{q}$ splitting of a soft gluon, where the q and $\bar{q}$ then enter separate jets. A simple modification [16] of the distance measure in the k_t algorithm can solve the problem and make the flavour IRC safe.

A key advantage of the resulting IRC safe "flavour-k_t" algorithm emerges when talking about *heavy*-flavour jet cross sections. With unsafe definitions, higher orders involve powers of large logarithms $\ln p_t/m_b$, giving large NLO scale uncertainties. This is especially the case for current experimental b-jet measurements, in which a jet containing a b and a $\bar{b}$ is considered to be a normal b-jet. With a proper, IRC safe definition most of the large logarithms disappear, and remaining ones can be absorbed into the parton distribution functions. The result is a reduction in the theory uncertainty for the inclusive b-jet spectrum from $\sim$ 40–50% (see *e.g.* [17]) to the $\sim$ 10–20% shown in Fig. 4 [15] (calculated with a modified version of `nlojet++` [18]). It

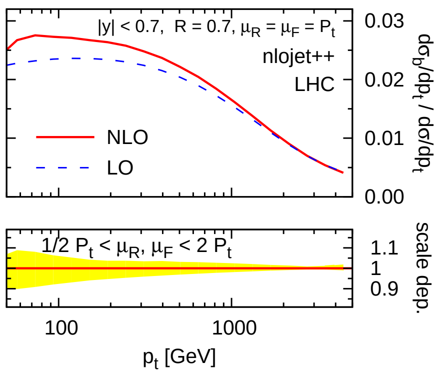

Fig. 4. b-jet fraction at LHC and scale dependence, with flavour-k_t jets [15].

[2] On the other hand, for hard particles, modern cone algorithms like SISCone have a reach that extends somewhat beyond R and this can lead to issues in resolving complex multi-jet events.

should be said that while normal IRC safe jet algorithms involve no particular experimental issues, the flavour-k_t algorithm does require a jet's flavour to be taken as the sum of the flavours of the jet's constituents — *i.e.* one should be able to distinguish a jet containing a b and $\bar{b}$ (not a b-jet) from one that contains just a single b. This is challenging experimentally, but if it can be done it will have significant benefits also in reducing QCD "b"-jet backgrounds (more than 50% of which come from $b\bar{b}$ jets) in new-physics searches. A measurement of the flavour-k_t b-jet spectrum would then provide a powerful cross-check that the separation of $b\bar{b}$ and b jets is being done effectively.

4. Outlook

Practical, infrared and collinear safe options now exist for both cone and sequential recombinations jet algorithms, and are in the process of being incorporated into the LHC experiments' software frameworks. If they are widely adopted (not to be taken for granted given the continued presence also of long-established unsafe options and the inertia inherent in large organisations), hadron-collider jet-finding will finally come into accord with the 1990 Snowmass principles, a key step if the LHC is to benefit from the ongoing and extensive calculational effort in QCD.

There is more, of course, to jet finding than just practicality and IRC safety. Most of the thought about jet algorithms, historically, has been for low-noise, single-scale, quark-jet dominated environments such as LEP. Work is now being carried out that addresses the significant novel issues that arise at the LHC. There is considerable scope for further work, and it is to be hoped that this, together with open exchanges between theorists and experimenters, will help us make the best possible use of jets in the rich environment of LHC.

I wish to thank Andrea Banfi, Matteo Cacciari, Mrinal Dasgupta, Lorenzo Magnea, Gregory Soyez and Giulia Zanderighi for stimulating collaborations on the topics described here, as well as Günther Dissertori, Steve Ellis, Joey Huston and Markus Wobisch for numerous interesting discussions. Work supported in part by contract ANR-05-JCJC-0046-01.

REFERENCES

[1] J.E. Huth *et al.*, Fermilab-Conf-90-249-E.
[2] N. Kidonakis, G. Oderda, G. Sterman, *Nucl. Phys.* **B525**, 299 (1998).
[3] G.C. Blazey *et al.*, hep-ex/0005012.

[4] G.P. Salam, G. Soyez, *J. High Energy Phys.* **0705**, 086 (2007).

[5] S. Catani, Y.L. Dokshitzer, M.H. Seymour, B.R. Webber, *Nucl. Phys.* **B406**, 187 (1993); S.D. Ellis, D.E. Soper, *Phys. Rev.* **D48**, 3160 (1993).

[6] M. Cacciari, G.P. Salam, *Phys. Lett.* **B641**, 57 (2006) [hep-ph/0512210], http://www.lpthe.jussieu.fr/ salam/fastjet/

[7] J.-D. Boissonnat *et al.*, *Comp. Geom.* **22**, 5 (2001).

[8] M. Cacciari, G.P. Salam, G. Soyez, LPTHE-07-02, in preparation.

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

[10] A. Abulencia *et al.* [CDF — Run II], *Phys. Rev.* **D75**, 092006 (2007).

[11] T. Sjöstrand *et al.*, *Comput. Phys. Commun.* **135**, 238 (2001); hep-ph/0308153.

[12] I.P. Lokhtin, A.M. Snigirev, *Eur. Phys. J.* **C45**, 211 (2006); *J. Phys.* **G34**, S999 (2007) [hep-ph/0612109].

[13] M. Dasgupta, L. Magnea, G.P. Salam, *J. High Energy Phys.* **0802**, 055 (2008) [arXiv:0712.3014 [hep-ph]].

[14] M. Dasgupta, G.P. Salam, *J. Phys. G* **30**, R143 (2004).

[15] A. Banfi, G.P. Salam, G. Zanderighi, *J. High Energy Phys.* **0707**, 026 (2007).

[16] A. Banfi, G.P. Salam, G. Zanderighi, *Eur. Phys. J.* **C47**, 113 (2006).

[17] CDF Collaboration, Note 8418; see also M.D'Onofrio [CDF and DØ Collaborations], FERMILAB-CONF-06-224-E.

[18] Z. Nagy, *Phys. Rev. Lett.* **88**, 122003 (2002); *Phys. Rev.* **D68**, 094002 (2003).