Published by Institute of Physics Publishing for SISSA

Received: *April 25, 2007*
Accepted: *June 28, 2007*
Published: *July 10, 2007*

Accurate QCD predictions for heavy-quark jets at the Tevatron and LHC

Andrea Banfi

Università degli Studi di Milano-Bicocca and INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy
E-mail: andrea.banfi@mib.infn.it

Gavin P. Salam

LPTHE, Université Pierre et Marie Curie – Paris 6, Université Denis Diderot – Paris 7, CNRS UMR 7589, 75252 Paris cedex 05, France
E-mail: salam@lpthe.jussieu.fr

Giulia Zanderighi

Theory Division, Physics Department, CERN, 1211 Geneva 23, Switzerland
E-mail: giulia.zanderighi@cern.ch

Abstract: Heavy-quark jets are important in many of today's collider studies and searches, yet predictions for them are subject to much larger uncertainties than for light jets. This is because of strong enhancements in higher orders from large logarithms, $\ln(p_t/m_Q)$. We propose a new definition of heavy-quark jets, which is free of final-state logarithms to all orders and such that all initial-state collinear logarithms can be resummed into the heavy-quark parton distributions. Heavy-jet spectra can then be calculated in the massless approximation, which is simpler than a massive calculation and reduces the theoretical uncertainties by a factor of three. This provides the first ever accurate predictions for inclusive b- and c-jets, and the latter have significant discriminatory power for the intrinsic charm content of the proton. The techniques introduced here could be used to obtain heavy-flavour jet results from existing massless next-to-leading order calculations for a wide range of processes. We also discuss the experimental applicability of our flavoured jet definition.

Keywords: QCD, Jets, NLO Computations, Hadronic Colliders.

JHEP07(2007)026

Contents

JHEP07(2007)026

1. Introduction

Studies of heavy-quark jets, i.e. charm and bottom jets, are important for a range of reasons. They are of intrinsic interest because charm and bottom are the flavours for which there exists the most direct correspondence between parton level production and the observed hadron level. They have the potential to provide information on the c- and b-quark parton distribution function (PDF), which are the only components of proton structure that are thought to be generated entirely perturbatively from the DGLAP evolution of the other flavours. Furthermore, b-jets enter in many collider searches, notably because they are produced in the decays of various heavy particles, e.g. top quarks, the Higgs boson (if light) and numerous particles appearing in proposed extensions of the Standard Model (SM) [1].

Within the SM a range of production channels exist for heavy-quark jets, e.g. pure QCD production or in association with heavy bosons ($W, Z, H \ldots$), see e.g. [2]. The simplest and most fundamental measurement of heavy-quark jet production is the inclusive heavy-quark jet spectrum, which is dominated by pure QCD contributions. Predictions for this sort of quantity have always been obtained using calculations in which the c or b quark has been explicitly taken to be massive while all other lighter masses are neglected.

An example is the inclusive b-jet spectrum measured by CDF [3]. Figure 1 shows the ratio of the experimental measurement to the next-to-leading order (NLO) calculation of [4]. A striking feature of this plot is the size of the theoretical (scale variation) uncertainties ($\sim 50\%$). One notes in particular that there is a significant region where the experimental uncertainties are smaller than the theoretical ones. Furthermore, the b-jet

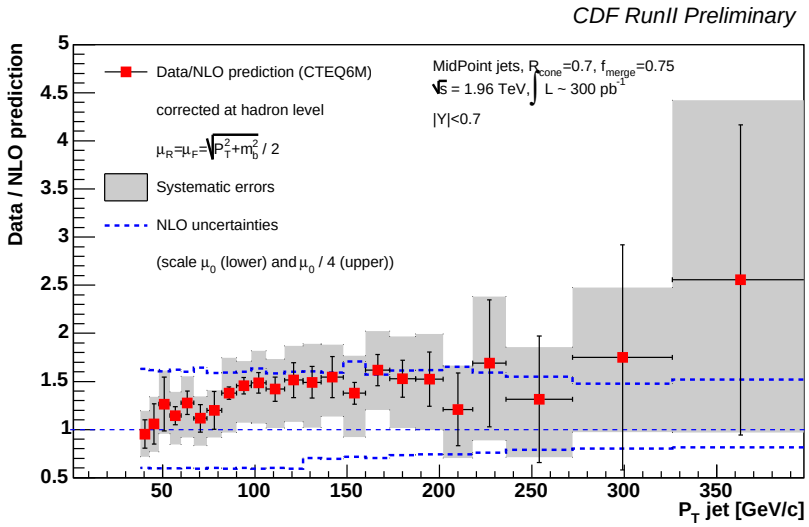

Figure 1: Ratio of the measured inclusive b-jet spectrum to NLO prediction. The measurement is performed for jets with transverse momentum $38\,\mathrm{GeV} < P_{T,\mathrm{jet}} < 400\,\mathrm{GeV}$ and rapidity $|y_{\mathrm{jet}}| < 0.7$. The plot is taken from ref. [3].

theory uncertainties are considerably larger than the corresponding ones for the normal (light) jet inclusive spectrum ($\sim$ 10-20%), see for example [5].

The origin of the large theoretical uncertainties in figure 1 can be understood by examining figure 2. Its top panels show the K-factor (NLO/LO) as obtained with MCFM for the Tevatron Run II ($p\bar{p}$, $\sqrt{s} = 1.96\,\mathrm{TeV}$, left) and for the LHC (pp, $\sqrt{s} = 14\,\mathrm{TeV}$, right).[1] The fact that the K-factor is considerably larger than one indicates that the perturbative series is very poorly convergent, and implies that the NLO result cannot be an accurate approximation to the full result. It is for this reason that the scale dependence (middle panels) is large. One might think that a calculation with MC@NLO [12] should do better, since it includes both NLO and all-order resummed logarithmically enhanced terms. This turns out not to be the case, as can be seen from its persistently large scale dependence.[2] Essentially, while MC@NLO contains a good matching between the NLO b-production calculation and the b-quark fragmentation logarithms in Herwig, it does not match with the logarithmic enhancements contained in Herwig for b-quark production, but rather just replaces them with the NLO result.

The poor convergence of the perturbative series is related to the different channels for heavy quark production. At leading order (LO), only the so-called flavour creation channel (FCR) is present, $\ell\ell \to b\bar{b}$, where ℓ is a generic light parton (quark or gluon). At NLO, two new channels open up, often referred to as the flavour excitation (FEX) and gluon

[1] Figure 1 has been obtained using a midpoint type [6] cone algorithm, however given the recent discoveries [7, 8] of infrared safety issues in midpoint cone algorithms, we prefer to illustrate our arguments with an inclusive k_t-algorithm [9]. In practice, we expect most features of the figure to be insensitive to the choice of algorithm, for example also with an infrared safe cone-type algorithm such as SISCone [8].

[2] Poor numerical convergence prevented us from presenting the scale dependence for MC@NLO at the LHC. Note also that no K-factor has been shown for MC@NLO because the LO result is not unambiguously defined.

JHEP07(2007)026

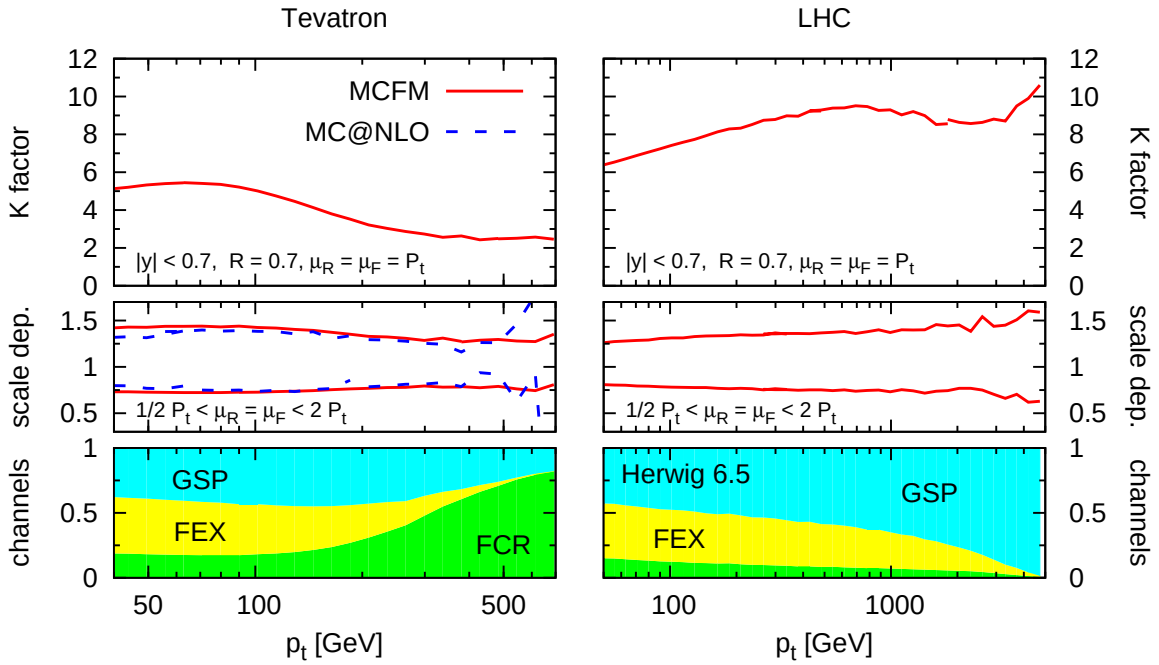

Figure 2: Top: K-factor for inclusive b-jet spectrum as computed with MCFM [10], clustering particles into jets using the k_t jet algorithm [9] with R=0.7, and selecting jets in the central rapidity region ($|y| < 0.7$). Middle: scale dependence obtained by simultaneously varying the renormalisation and factorisation scales by a factor two around P_t, the transverse momentum of the hardest jet in the event. Bottom: breakdown of the Herwig [11] inclusive b-jet spectrum into the three major hard underlying channels cross sections (for simplicity the small $bb \to bb$ is not shown).

splitting channels (GSP).[3] In the former, a gluon from one of the incoming hadrons splits collinearly into a $b\bar{b}$-pair and one of those b-quarks enters the hard $b\ell \to b\ell$ scattering. In the gluon splitting process, the hard scattering process is of the form $\ell\ell \to \ell\ell$, and one of the final-state light partons (at NLO always a gluon) splits collinearly into a $b\bar{b}$-pair (a jet containing both b and $\bar{b}$ is considered to be a b-jet in standard definitions). The various channels can be conveniently separated with a parton shower Monte Carlo generator such as Herwig [11], where one can determine the underlying hard channel from the hard process in the Herwig event record.[4] Their relative contributions to the total b-jet spectrum are

[3]It is sometimes stated that it makes no sense, beyond LO, to separately discuss the different channels, for example because diagrams for separate channels interfere. However, each channel is associated with a different structure of logarithmic enhancements, $\ln^n(p_t/m_b)$, and so there is distinct physical meaning associated with each channel. Furthermore one can give a precise, measurable definition to each channel, e.g. using an exclusive variant of the flavour jet algorithm discussed below. Though there will be some arbitrariness in any such definition, relating to the choice of parameters of the jet algorithm, this arbitrariness is no more troubling conceptually or practically than the jet-definition dependence that arises when determining the number of jets in an event.

[4]The use of Herwig to label the flavour channel does not correspond to a directly measurable definition, however Herwig does have the correct logarithmic enhancements for each channel, and furthermore gives results rather similar to those based on a flavour-channel classification using the algorithm of section 2 with $R = 1$ (the value of R that places initial and final-state radiation on the same footing for k_t type jet

JHEP07(2007)026

shown in the bottom panel of figure 2. One sees that the supposedly LO channel (FCR) is nearly always smaller than the two channels that at fixed order enter only at NLO (FEX and GSP). This is because both NLO channels receive a strong enhancement from collinear logarithms, going as $\alpha_s^2(\alpha_s \ln(p_t/m_b))^n$ for flavour excitation [13] and $\alpha_s^2 \cdot \alpha_s^n \ln^{2n-1}(p_t/m_b)$ for gluon splitting ($n \geq 1$) [14].

Three approaches come to mind for increasing the accuracy of the b-jet spectrum prediction. The most obvious (and hardest) is to carry out the full massive next-to-next-to-leading order calculation. Aside from being beyond the limit of today's technology, such an approach would still leave many of the higher order logarithms uncalculated and so would only partially improve the situation. A second approach would be to carry out the explicit resummation of both the incoming and outgoing collinear logarithms. The technology for each resummation on its own is well-known at next-to-leading logarithmic accuracy (NLLA) [13 – 15], though significant effort would probably be necessary to assemble them together effectively. In both of the above approaches, the largest residual uncertainties are likely to be associated with the channel with the most logarithms, gluon splitting. This channel however does not even correspond to one's physical idea of a b-jet, i.e. one induced by a hard b-quark and it seems somehow unnatural to include it at all as part of one's b-jet spectrum.

We therefore propose a third approach to improving the accuracy of the prediction of the b-jet spectrum. It is a two-pronged approach. Firstly, one uses a definition of b-jets which maintains the correspondence between partonic flavour and jet flavour. Specifically, we take the flavour-k_t algorithm of [16]. Within this algorithm, described in section 2, a jet containing equal number of b quarks and b antiquarks is considered to be a light jet, so that jets that contain a b and $\bar{b}$ from the gluon splitting channel do not contribute to the b-jet spectrum. The use of this kind of algorithm already leads to some reduction of the theoretical uncertainty on the b-jet spectrum with a standard massive calculation (e.g. with MCFM). Further improvement can be obtained by exploiting the fact that the logarithms of p_t/m_b that remain are those associated with flavour excitation, which coincide with those resummed in the b-quark parton distribution function (PDF) at scale p_t. If one uses a b-quark PDF to resum these logarithms, no other logarithms $\ln(p_t/m_b)$ appear in the rest of the calculation, so that one can safely take the limit $m_b \to 0$ and one misses only corrections suppressed by powers of $(m_b/p_t)^2$, possibly with additional logarithms.[5] The validity of this procedure is a consequence of the infrared safety of the jet-flavour even in the massless

algorithms).

[5]Specifically, at leading order we expect pure m_b^2/p_t^2 corrections. Beyond LO we expect such corrections to be enhanced by logarithms similar to those that arise for the b-quark multiplicity [14], yielding terms of relative order $\alpha_s^n\,(m_b^2/p_t^2)\,\ln^{2n-1}(p_t/m_b)$. Starting from NNLO one also becomes sensitive to a kinematic region in which a pair of b-quarks are produced from the large-angle splitting of a soft gluon with $p_t \sim m_b$. Depending on the recombination scheme, this can lead to finite-mass effects being suppressed only by m_b/p_t (much as was observed for hadron-mass effects in [17]), and at all orders one expects them to then contribute terms $\alpha_s^{2+n}(m_b/p_t)\ln^n(p_t/m_b)$ with resummations similar to those discussed in [15, 17]. We note that all such effects are relevant also in normal inclusive jet calculations, since there too kinematic effects associated with b masses are neglected.

limit (see later).[6] This third approach is therefore the one which is technically the easiest to pursue and which should simultaneously reduce the theoretical uncertainties the most. In section 3 we present results for c- and b-jets using this method. A similar approach can be used also in different contexts, e.g. recently the flavour-algorithm of [16] has been used to define the e^+e^- forward-backward asymmetry for b in an infrared-safe way, making it possible to compute this quantity at NNLO using a massless QCD calculation [19].

Several issues deserve detailed discussion in the above approach. Firstly for moderate values of p_t (or of the jet energy in e^+e^-), finite-mass effects may not be completely negligible. It is therefore important to determine their size. We explain briefly how this can be done in section 3, with further details given in appendix A. A second issue is an experimental one related to the limited efficiency for the identification of B and D hadrons. Though not strictly within the remit of a theoretical paper, we do find it useful to discuss various points related to this issue in section 4. Finally we also comment on the question of electroweak effects, in appendix B.

2. The heavy-quark jet algorithm

In general, flavour-algorithms provide an IR-safe definition of the flavour of a jet, provided one knows the (light or heavy) flavour of each parton involved. However to study heavy-quark jets it is not necessary to know the flavour of light quarks, because gluons and light flavoured quarks can be considered as flavourless, while one assigns to heavy (anti)-quarks a flavour 1 (-1). We define the heavy-flavour of a (pseudo)-particle or a jet as its net heavy flavour content, i.e. the total number of heavy quarks minus heavy antiquarks. One may alternatively use the sum of the number of quarks and anti-quarks modulo 2. Flavourless (flavoured) objects are then those with (non-)zero net flavour. We present here the inclusive version of the heavy-flavour jet algorithm for hadron-hadron collisions, referring the reader to [16] for the motivation of the formalism (as well as the original exclusive formulation):

1. For any pair of final-state particles i, j define a class of longitudinal boost invariant distances $d_{ij}^{(F,\alpha)}$ parametrised by $0 < \alpha \leq 2$ and a jet radius R

$$d_{ij}^{(F,\alpha)} = \frac{\Delta y_{ij}^2 + \Delta\phi_{ij}^2}{R^2} \times \begin{cases} \max(k_{ti}, k_{tj})^\alpha \min(k_{ti}, k_{tj})^{2-\alpha}\,, & \text{softer of } i,j \text{ is flavoured,} \\ \min(k_{ti}^2, k_{tj}^2)\,, & \text{softer of } i,j \text{ is flavourless,} \end{cases} \tag{2.1}$$

   where $\Delta y_{ij} = y_i - y_j$, $\Delta\phi_{ij} = \phi_i - \phi_j$ and k_{ti}, y_i and ϕ_i are respectively the transverse momentum, rapidity and azimuth of particle i, with respect to the beam.

   For each particle define a distance with respect to the beam B at positive rapidity,

$$d_{iB}^{(F,\alpha)} = \begin{cases} \max(k_{ti}, k_{tB}(y_i))^\alpha \min(k_{ti}, k_{tB}(y_i))^{2-\alpha}\,, & i \text{ is flavoured,} \\ \min(k_{ti}^2, k_{tB}^2(y_i))\,, & i \text{ is flavourless,} \end{cases} \tag{2.2}$$

[6] We note that such a '5-flavour scheme', with resummed b-quark PDFs has been used before in MCFM for $H + b$ and $Z + b$ production [18]. In that case, because a non-flavour jet algorithm was used, it was necessary to supplement the results with an explicit massive calculation of the NLO gluon-splitting process. We thank John Campbell for bringing this to our attention.

JHEP07(2007)026

with

$$k_{tB}(y) = \sum_i k_{ti} \left(\Theta(y_i - y) + \Theta(y - y_i) e^{y_i - y} \right) . \tag{2.3}$$

Similarly define a distance to the beam $\bar{B}$ at negative rapidity by replacing k_{tB} in eq. (2.2) with $k_{t\bar{B}}$

$$k_{t\bar{B}}(y) = \sum_i k_{ti} \left(\Theta(y - y_i) + \Theta(y_i - y) e^{y - y_i} \right) . \tag{2.4}$$

2. Identify the smallest of the distance measures. If it is a $d_{ij}^{(F,\alpha)}$, recombine i and j into a new particle, summing their flavours and 4-momenta; if it is a $d_{iB}^{(F,\alpha)}$ (or $d_{i\bar{B}}^{(F,\alpha)}$) declare i to be a jet and remove it from the list of particles.

3. Repeat the procedure until no particles are left.

Sensible values for α are 1 or 2 [16] and R should both be kept of order 1, to avoid the appearance of large logarithms of R.

The IR-safety of this algorithm was proved in [16]. A general consequence of IR-safety is that it allows one to take the limit $m_Q^2 \to 0$ (any finite-mass corrections being suppressed by powers of $m_Q^2/{p_t}^2$) as long as collinear singularities associated with incoming heavy quarks are factorised into a heavy quark PDF. This means that we can compute heavy-quark jet cross sections using a simpler, light-flavour NLO program, rather than a heavy-flavour one [20]. Furthermore IR and collinear safety ensure that one obtains the same results whether one considers heavy-quark flavour at parton level, or heavy-meson flavour at hadron level, modulo corrections suppressed by powers of Λ_{QCD}/p_t.

3. Results

In figure 3 we present the inclusive b-jet p_t-spectrum as obtained with the flavour algorithm specified above. We have used the jet-algorithm parameters $\alpha = 1$, and $R = 0.7$, the latter having been shown to limit corrections associated with the non-perturbative underlying event [5]. The left (right) column of the figure shows results for the Tevatron run II (LHC). We have selected only those jets with rapidity $|y| < 0.7$. We also show the full inclusive jet spectrum (all jets) as obtained with a standard inclusive k_t-algorithm with $R = 0.7$.

The spectra have been calculated using NLOJET [21]. The publicly available version sums over the flavour of outgoing partons. We therefore had to extend it so as to have access to the flavour of both incoming and outgoing partons. We fixed the default renormalisation and the factorisation scales to be P_t, the transverse momentum of the hardest jet in the event and chose as a default PDF set CTEQ61m [22]. We also used the a posteriori PDF library (APPL) of [23], together with the HOPPET [24] and LHAPDF [25] packages to allow us to vary scales and PDF sets after the NLOJET Monte Carlo integration.

The figure shows the inclusive jet spectrum at LO (blue, dashed) and at NLO (red, solid) for all jets and for b-jets. The b-jet cross section is always a few percent of the

JHEP07(2007)026

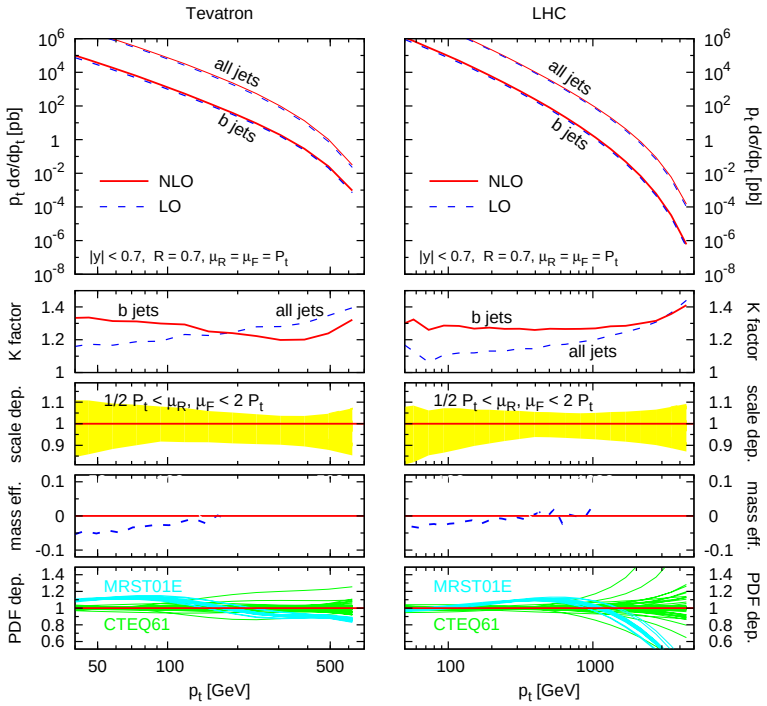

Figure 3: Inclusive jet spectrum at the Tevatron (left) and at the LHC (right). The top two panels show results for both b-jets and all-jets, while the lower three panels apply only to b-jets. See text for further details.

light jet one. The K-factor, the ratio of NLO over LO cross-section is shown below and is similar (between 1.15 and 1.4) for light and b-jets, both at the Tevatron and at the LHC. To provide an estimate for the theoretical uncertainty we vary separately the factorisation and the renormalisation scale in the range $1/2P_t < \mu_R, \mu_F < 2P_t$. The band associated with this variation is shown in the plots below. We see that this is at most a 15% effect in the region considered. We note that our procedure is more conservative than the usual simultaneous variation of μ_R and μ_F (as done in figures 1 and 2).

We have also calculated (but do not show) the b-jet spectrum for our definition of heavy jets using a massive NLO calculation with MCFM [10]. We find that the results are consistent with those from the massless calculation, though the uncertainties in the massive calculation are much larger, only slightly smaller than those in figure 2.

Though the massive calculation is not itself of much direct interest given its significant uncertainties, it does enable one to estimate residual finite-mass effects, via the relation

$$\left.\frac{d\sigma_b}{dp_t}\right|_{m=m_b} = \frac{d\sigma_b^{\text{nlojet}}}{dp_t} + \lim_{m_0\to 0}\left(\left.\frac{d\sigma_b^{\text{MCFM}}}{dp_t}\right|_{m=m_b} - \left.\frac{d\sigma_b^{\text{MCFM}}}{dp_t}\right|_{m=m_0} + C(p_t)\ln\frac{m_b}{m_0}\right). \quad (3.1)$$

Here, the contents of the bracket corresponds to the evaluation of the difference between the result for the true mass and the massless limit, while subtracting logarithms such that the massless MCFM calculation is effectively being carried out with a coupling and b-PDF that are mass-independent at scale p_t. Further details and the form for $C(p_t)$ are provided in appendix A. The relative size of the residual finite mass effects is shown in

JHEP07(2007)026

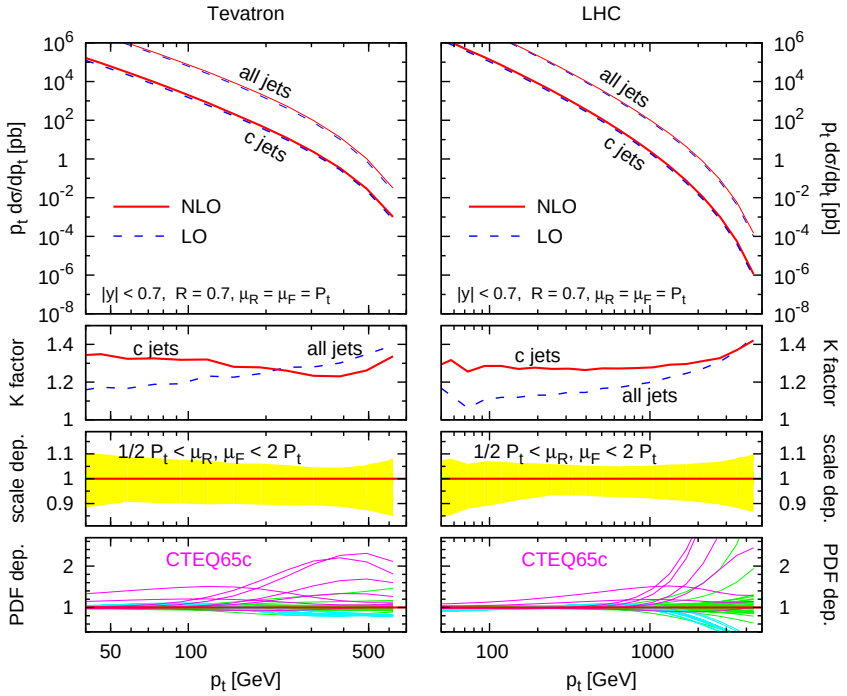

Figure 4: Inclusive jet spectrum at the Tevatron (left) and at the LHC (right) for generic jet production and for c-jet production. See text for further details.

the penultimate panel of figure 3. They decay somewhat more slowly with p_t than the naive expectation of m_b^2/p_t^2 (a feature noted before in [26]), perhaps because they have logarithmic enhancements, cf. footnote 5. Nevertheless they are always below 6% and given their modest size compared to the massless perturbative uncertainties, we choose not to explicitly add them to the main NLOJET results.

To illustrate the dependence on the parton densities we show in the bottom panel of figure 3 the effect of using all members of the CTEQ61 [22] and MRST2001E [27] PDF sets, relative to the default CTEQ61m choice.[7] We see that the effect is always moderate at the Tevatron ($\lesssim 20\%$), while it is large at the LHC in the high p_t region, presumably because the b and gluon PDFs are not well constrained in that region.

We have also calculated the spectrum for charm jets and the results are shown in figure 4. We omit the panel showing finite-mass effects because of the low charm quark mass. The most notable difference relates to the PDF dependence. There has been some discussion of a possible intrinsic charm (IC) component of the proton and a recent analysis provides PDF sets, CTEQ65c, with various models for such a component, see [30] and references therein. One sees that at moderate p_t these sets suggest that there is up to 40% uncertainty in the charm jet spectrum and at higher p_t the uncertainty reaches a factor two. Further investigation reveals that the moderate p_t uncertainty is related to a possible sea-like IC component. In the sea-like scenario considered in [30] it was assumed that charm and anti-charm are distributed as the up and down sea components in the

[7] We have also examined the CTEQ65 [28], MRST2004nlo and MRST2004nnlo [29] sets and found similar results.

JHEP07(2007)026

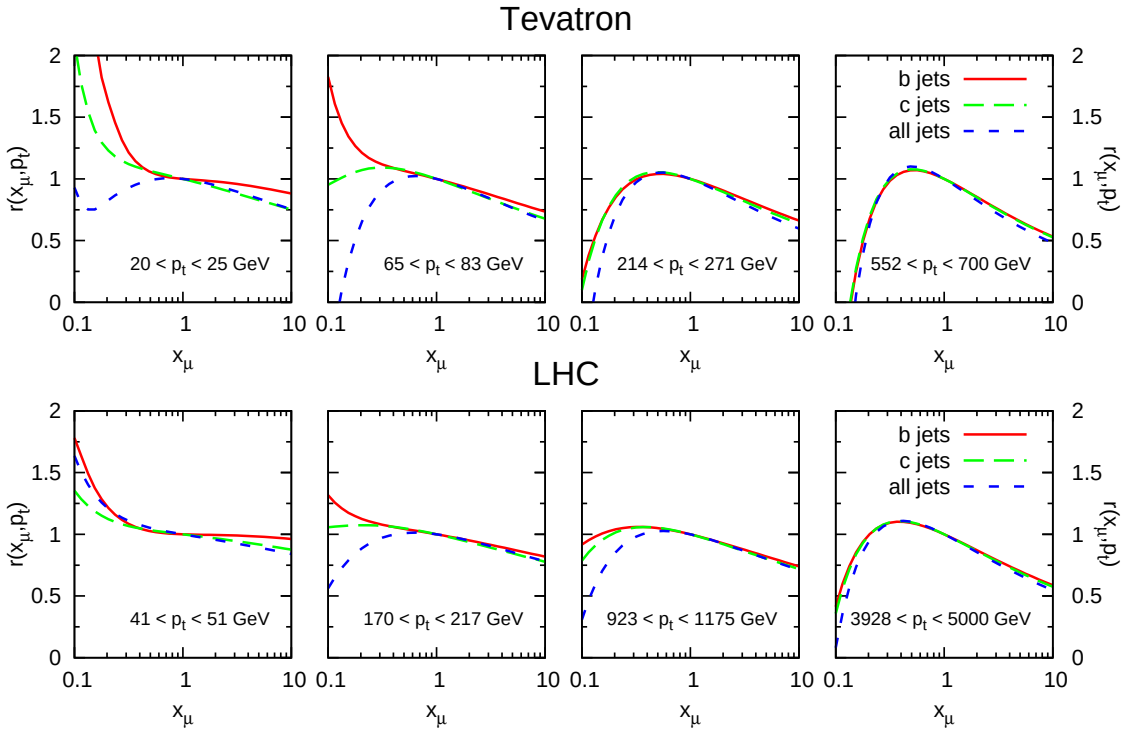

Figure 5: Ratio of spectrum at factorisation and renormalisation scale $\mu_R = \mu_F = x_\mu P_t$ and at $\mu_R = \mu_F = P_t$.

proton. At higher p_t the uncertainty is due to the valence-type models for IC considered in [30], specifically the original BHPS light-cone model [31], and a meson-cloud picture [32] in which the IC arises from virtual low-mass meson+baryon components of the proton.

Let us now return to the question of theoretical uncertainties in our predictions, specifically the scale dependence. Figure 5 shows the ratio

$$r(x_\mu, p_t) = \frac{\frac{d\sigma}{dp_t}(\mu = x_\mu P_t)}{\frac{d\sigma}{dp_t}(\mu = P_t)}, \tag{3.2}$$

for the inclusive and heavy-quark jet cross-sections in various p_t-bins. The factorisation and renormalisation scales are varied simultaneously, $\mu_R = \mu_F = \mu = x_\mu P_t$. At low p_t at the Tevatron and at intermediate p_t at the LHC the scale dependences are quite different at low values of x_μ ($\lesssim 0.5$) due to the dominance of different partonic channels. However, the sensitivity, i.e. the dependence of $r(x_\mu, p_t)$ on x_μ remains always of the same order for heavy-quark jets and all jets. The charm ratio is generally intermediate between the b and all-jet ratios, as is natural given the relative masses of the charm and bottom quarks.

The fact that the scale dependences are similar for all and heavy jets in many of the p_t bins, suggests that if one considers the ratio of heavy to all jets a significant part of the theory uncertainties may cancel. Additionally, a number of experimental uncertainties may cancel, for example part of the jet energy scale and luminosity dependence.

Accordingly in figure 6 we show the ratio of b- and c-jet spectra to the all-jet spectra. The ratio is always of the order of a few percent and is somewhat larger for c-jets than

JHEP07(2007)026

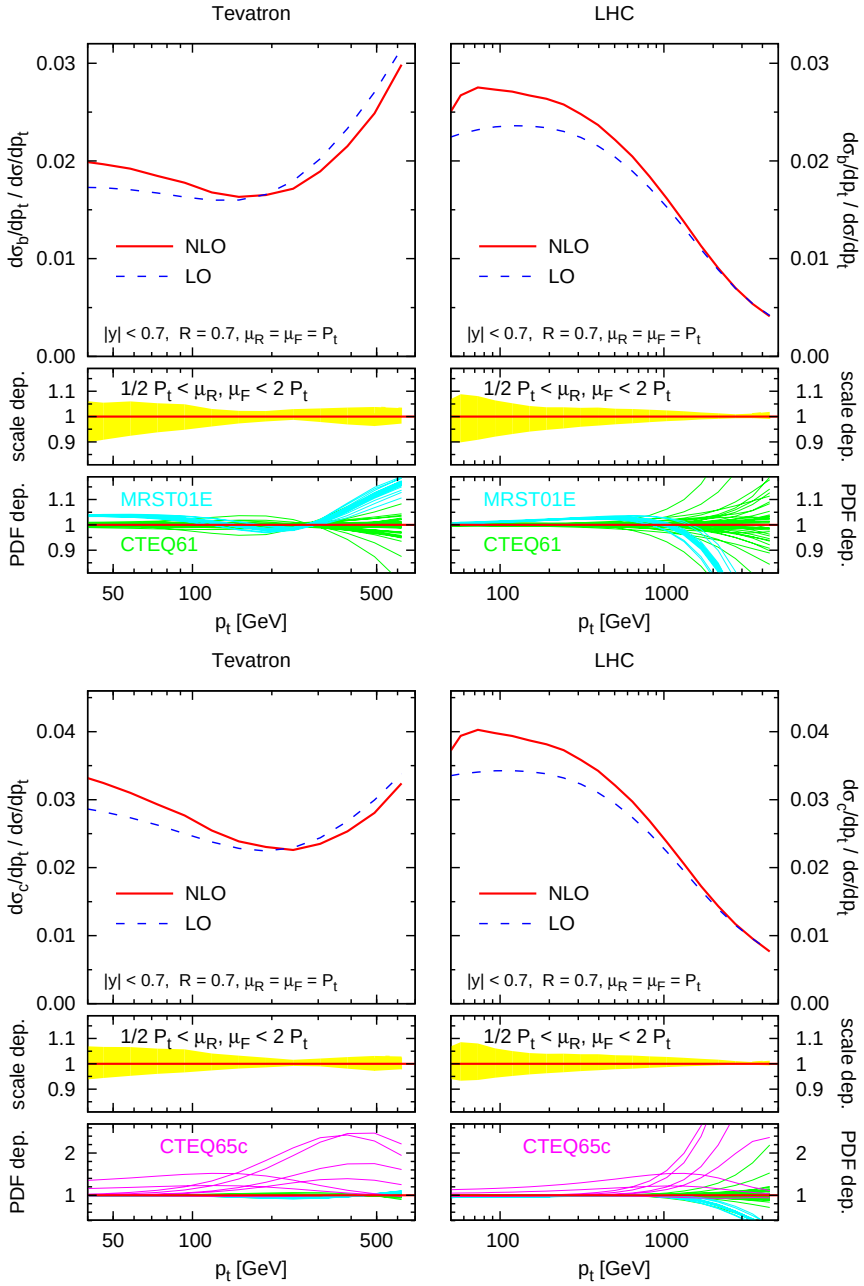

Figure 6: Top: ratio of b-jet to inclusive jet spectra at the Tevatron and at the LHC. Bottom: ratio of the c-jet to inclusive jet spectra. Further details are provided in the text.

for b-jets, as is to be expected given the larger charm PDF. At higher p_t it increases at the Tevatron and decreases at the LHC due to the different behaviour of the PDFs in the range of x and Q^2 probed by the two different machines. In particular, at large x all-jet spectra are dominated by channels with valence incoming quarks. The same is true at the Tevatron for heavy-quark jets, where the main high-p_t production channel is $q\bar{q} \to Q\bar{Q}$. At the LHC, on the contrary, high-p_t heavy quarks are produced mainly via $Qq \to Qq$ processes, so that heavy-jet spectra are suppressed by the heavy-quark PDF.

JHEP07(2007)026

The lower panels of figure 6 show the uncertainty associated with the variation of factorisation and renormalisation scales and the PDFs. The scale dependence is reduced in the whole p_t range compared to that for the heavy-jet spectra. This is especially the case at large p_t (cf. figure 5). The PDF dependence is also reduced except in the case of charm jets using PDFs with an intrinsic charm component, CTEQ65c.

4. Experimental issues

The main outstanding question is that of the experimental measurability of heavy flavour jets as defined here. We examine this specifically for b-jets, since they have been much more widely studied. We will comment briefly on c-jets at the end of the section.

The question of the experimental measurability of b-jet definition can only truly be settled by a detailed experimental study. However several points lead us to believe that such a measurement might well be possible. Our discussion here is inspired in part by that in [33], which measured $B\bar{B}$ azimuthal correlations at the Tevatron, including the region of small angular separation between the B and $\bar{B}$, which is the experimentally non-trivial region also for our definition of b-jets. One should be aware in the discussion below that the correspondence between our needs and what was done in [33] is only partial, insofar as the measured B-hadrons were not used as inputs to a jet algorithm, and also had lower typical transverse momenta than would the B-hadrons in b-jet studies.[8]

In an ideal world the input to the jet flavour algorithm would be a list of momenta of all particles in the event together with information about which particles correspond to a B-hadron. We are allowed to use B-hadrons rather than b quarks in the algorithm because the flavoured jets are infrared and collinear safe and the fragmentation of a b quark into a B-hadron should have no more effect than collinear radiation from the b quark.

Experimentally one has information on charged tracks and their momenta, calorimeter energy deposits, and b-tags. The latter typically exploit the long lifetime of B-hadrons, which causes the B-hadron to decay some small but measurable distance away from the primary interaction vertex of the event. If the B-hadron decay products include two or more charged particles then a secondary vertex may be identified from the intersection of the resulting charged tracks, whereas if the decay involves only one charged track then one may still obtain a b-tag based on the finite impact parameter between that track and the primary vertex. Often the b-tagging is restricted to tracks within hard jets, so as to reduce certain backgrounds.

Current b-tagging abilities don't correspond to our 'ideal world' scenario for a variety of reasons. Firstly, since one often sees only a subset of the B-hadron decay products, one does not know the B-hadron momentum. This should not matter since the jet algorithm will in its first steps recombine the observed charged tracks in the decay with the calorimeter energy deposits from the neutral particles (other than neutrinos) in the decay.

[8] At higher p_t's the fraction of $b\bar{b}$ pairs at small angles will be increased, making the analysis more difficult, on the other hand the secondary vertices will be displaced further from the primary vertex and this should facilitate the analysis. The extent to which these two effects cancel can only be determined by a full experimental study.

A second problem is that whereas experiments first search for jets and then do the b-tagging, we need the information on b-tags before running the jet algorithm. This should not be a major obstacle: one may first identify jets using a standard k_t or cone algorithm, with large radius parameter (so as to catch most b's, as done in [33]), carry out the b-tagging, and then run the flavour algorithm using that information.

The third and potentially most serious issue relates to the finite efficiency for b-tagging, and notably for double b-tagging inside a single jet. The efficiency for b-tagging is limited for various reasons: partly because of the need to place cuts on impact parameter to avoid backgrounds from decays of charm hadrons, which also decay a small but measurable distance from their production vertex (such backgrounds are partially reduced also by using the invariant mass of the decay products); and partly because of issues related to detector limitations. Double b-tagging for a pair of B-hadrons that are close in rapidity and azimuth (i.e. in the same jet) is considered particularly difficult, because of the need to be sure that, if one sees two secondary vertices in a jet, they aren't 'sequential tags' from a single b, i.e. the vertex from a B-hadron decaying to a D-hadron plus other particles, followed by the vertex from the D decay. Double b-tagging inside a single jet is nevertheless possible, albeit currently with limited efficiency, as has been shown in [33]. This is important because our algorithm relies on jets with two b's inside being identified as light jets.

To evaluate the impact of finite efficiencies, we consider the following simple model. We suppose the efficiency for tagging a single B-hadron to be x, and the efficiency for tagging two B-hadrons that are well-separated (i.e. in separate jets) to be x^2. Typical values for x are ~ 0.5. In contrast the probability of tagging two nearby B-hadrons is taken to be yx^2 (while the probability for tagging neither is $(1-x)^2$), with $y \simeq 0.2$ [33] a measure of the extra difficulty of tagging two nearby B-hadrons. If, in a given bin, the number of true b-jets is T and the number of jets containing $b\bar{b}$ due to gluon splitting is G, then the measured number of single-tagged b-jets will be[9]

$$t = xT + x(2-(1+y)x)G\,. \tag{4.1}$$

The contamination due to single-tagged gluon splitting is found by taking one minus the fraction of gluon-splitting jets where neither b has been tagged, or where both b's have been tagged, $x(2-(1+y)x)G = (1-(1-x)^2 - x^2y)G$. The measured number of light, 'gluon-splitting', jets with double b tags will be

$$g = x^2 y\, G\,. \tag{4.2}$$

It is straightforward to deduce T from measurements of t and g,

$$T = \frac{t}{x} - \frac{2-(1+y)x}{x^2 y} g\,, \tag{4.3}$$

[9]We ignore the potential effect of a flavour-mistag on the kinematics of the jets. This should be justified since the differences between a flavour k_t and a normal k_t algorithm are at the level of a few percent in the spectra, and in the absence of flavour information the flavour k_t algorithm just behaves like a normal k_t algorithm.

JHEP07(2007)026

as long as one knows the efficiencies x and y. In practice those efficiencies will be imperfectly known, with uncertainties δx and δy, and the effect of the estimated efficiency, used in eq. (4.3), being different from the true efficiency, eqs. (4.1), (4.2), will be an error δT on the determination of T,

$$\delta T^2 = \left[(2G - T)\frac{\delta x}{x}\right]^2 + \left[G(2 - x)\frac{\delta y}{y}\right]^2 , \tag{4.4}$$

where we assume the uncertainties on x and y to be uncorrelated. Since G and T are of the same order of magnitude (cf. figure 2), the uncertainty on T is essentially given by the relative uncertainties on x and y. If these can both be controlled to within 10%[10] then for $G \simeq 0.75\,T$, as we have at the Tevatron for $p_t \sim 100\,\mathrm{GeV}$, the relative uncertainty on T should be roughly 12% (for $x \simeq 0.5$). For an integrated luminosity of $2\,\mathrm{fb}^{-1}$ there are $\sim 10^5$ events in a bin of width 10 GeV centred at $p_t = 100\,\mathrm{GeV}$, so statistical errors will be considerably smaller than this, and they are dominated by the relative error on $g = x^2 y G$. Only at higher energies, when g starts to be small, will the enhancement of relative statistical errors due to the limited tagging efficiencies start to matter.

The above discussion is of course somewhat simplistic. In reality, single and double b-tagging efficiencies may vary with rapidity, azimuthal separation and transverse momentum, though this ought to be possible to account for; one should also correct for impurities in the b-tag samples — based on the uncertainties for the azimuthal correlations given in [33], this may be roughly equivalent to doubling the uncertainty on y; and a number of other experimental uncertainties will also contribute, such as energy scale uncertainties. On the other hand, steady progress is being made in b-tagging techniques [34 – 36]. One also wonders whether the knowledge that a second b is present *somewhere* in the event can be used in conjunction with a loose second b-tag, so as to obtain information about where the second b is most likely to be (in the same jet, in another jet, or down the beam-pipe), giving an effectively larger value for y (possibly even > 1). This might be important particularly when statistics are limited, e.g. at high p_t and also potentially when using flavour information in new-particle searches.

Finally, as concerns c-jets, though they have been the subject of far fewer investigations, we do note that double-tag samples also exist for charmed hadrons [37] and that some of the studies on b-tagging [34] also provide information on charm flavour, suggesting that c-jet studies may also be possible. As for b-jets, a critical issue in a good measurement of the charm jet spectrum will be not so much that of obtaining high tagging efficiencies, but rather of a good understanding of those efficiencies even if they are low.

5. Conclusions

The key finding of this article is that if one uses a properly defined jet-flavour algorithm and exploits its infrared safety to take the massless limit, predictions for heavy-quark jet

[10] The most delicate is y, and from table III of [33], which contains a breakdown of sources of systematic error (including that on the relative efficiency for tagging two nearby b's compared to two well separated b's), it seems that 10% is a reasonable value for the uncertainty on y.

JHEP07(2007)026

spectra can be made substantially more accurate than those based on current definitions and NLO massive calculations (e.g. MNR [20], MCFM [10] or MC@NLO [12]). When quantified in terms of scale dependence, the QCD theoretical uncertainty is reduced from $30-50\%$ to $10-20\%$. This is because large higher-order logarithms that first appear at NLO in the massive calculation are either cancelled by the jet definition itself, or else absorbed into the heavy-quark PDF in such a way as to become part of the leading order contribution, so that the NLO term is truly a perturbative correction.

Measurements of the heavy-flavour jet spectra as presented here would be of interest for a range of reasons. Heavy-flavour jet spectra measured so far do not distinguish between 'true' heavy-flavour jets and gluon jets that fragment to $Q\bar{Q}$. Our definition instead provides just the true flavoured-jet component. Thus for the first time not only is the momentum of a hard parton a meaningful observable quantity (as *defined* by the jet algorithm), but so is its flavour.

More generally, heavy-flavour jets, in particular b-jets, are used in a variety of contexts, including PDF measurements, top quark studies, and searches for new particles. These can only benefit from a properly defined jet flavour. One example seen in section 3 is for the charm PDF: current measurements leave considerable room for a non-perturbative 'intrinsic' charm component in the proton, and given an experimental accuracy that matched the theoretical accuracy of our charm-jet predictions, significant constraints could be placed on this intrinsic component. Similarly, a measurement of $W+c$-jet production could help constrain the strange quark PDF [38].

To calculate the heavy-flavour jet spectra shown here, we used NLOJET. By default it sums over the flavours of outgoing partons, so we modified it so as to be able to disentangle the flavour information. Though not completely trivial, this was quite a bit simpler than writing a new NLO Monte Carlo program for a massless process, and very much simpler than writing the corresponding heavy-flavour Monte Carlo program. One could analogously extract the flavour information from the many other NLO Monte Carlo programs involving massless QCD particles, thus providing heavy-flavour jet predictions in a range of processes. The usefulness of the flavour information is such that we strongly encourage NLO (and NNLO) Monte Carlo authors to provide it by default.[11]

To supplement the massless calculation, we also investigated residual effects associated with the finite value of the b-quark mass. For jets with $p_t \gtrsim 50\,\mathrm{GeV}$ they were of the order of 5%, falling off rapidly at higher p_t. This was the most laborious part of our study, however given the small size of the effects we believe that it should be safe to neglect them in future NLO calculations of heavy-flavour jets for other processes. Only when considering NNLO heavy-flavour jet predictions, or low values of p_t at NLO, should it become mandatory to include finite mass effects.

The main open question remains that of the experimental usability of our jet-flavour

[11] That usefulness extends beyond the framework of the jet-flavour type algorithm used here. For example to improve the prediction for the current experimental definition of b-jets, one could use the prediction given here as a starting point and supplement it with an NLO $(\alpha_s^3+\alpha_s^4)$ calculation of the difference between the experimental definition and ours, which starts only at $\mathcal{O}\left(\alpha_s^3\right)$. In principle, given the recent NLO calculation of the $Q\bar{Q}$+jet cross section [39], the technology already exists for such a combination.

JHEP07(2007)026

algorithm, mainly because of its reliance on the correct identification of situations where a jet contains both a B (D) and a $\bar{B}$ ($\bar{D}$) hadron. As discussed in section 4, given reasonable relative uncertainties on single and double-tag efficiencies, we believe that it ought to be possible to make an experimental measurement with errors that are not disproportionate compared to theory uncertainties. For the case of B hadrons, ongoing improvements in flavour tagging techniques, together with the use of 'loose' tagging to identify the second B hadron in an event where a first B hadron has already been found, might help further. We look forward therefore to future experimental investigations of heavy-flavour jet spectra with the definition presented here.

Acknowledgments

We are thankful to Matteo Cacciari, Mario Campanelli, Monica D'Onofrio, Stefano Frixione, Michelangelo Mangano, Andrea Rizzi, Ariel Schwartzman, Sofia Vallecorsa and Bryan Webber for fruitful discussions. We also thank John Campbell for assistance with MCFM. GZ would like to thank Zürich University for hospitality and the use of computer facilities. This work was supported in part by grant ANR-05-JCJC-0046-01 from the French Agence Nationale de la Recherche.

A. Finite mass effects

Given the small theoretical errors in the predictions for heavy-quark spectra when an infrared safe algorithm and massless calculation are used, it is important to make sure that the error due to the massless quark approximation remains smaller than the quoted theoretical errors even at moderate values of p_t. In this appendix we explain how $\mathcal{O}\left(\alpha_s^2\right)$ and $\mathcal{O}\left(\alpha_s^3\right)$ finite-mass effects can be extracted from the massive NLO calculation in MCFM.

The procedure consists in subtracting from the full, massive NLO result the collinear logarithms which with a massless calculation are resummed into heavy-quark PDFs, any remainder being due to finite mass effects $\mathcal{O}\left(m_Q^2/{p_t}^2\right)$ potentially enhanced by logarithms. The heavy-quark production mechanisms that can give rise to collinear logarithms are flavour excitation and gluon splitting. However, if an infrared safe algorithm is used the only logarithmic enhancements that survive are those associated with flavour excitation.

We denote generally by $\sigma(m_Q)$ any heavy-quark jet cross section corresponding to a set of kinematic cuts and study its dependence on the heavy-quark mass m_Q by considering $\Delta\sigma(m_Q, m_0) = \sigma(m_Q) - \sigma(m_0)$, where m_0 is an arbitrary reference mass. At NLO one can write $\Delta\sigma(m_Q, m_0) = \Delta\sigma_{\mathrm{L}}(m_Q, m_0) + \Delta\sigma_{\mathrm{R}}(m_Q, m_0)$, where $\Delta\sigma_{\mathrm{L}}(m_Q, m_0)$ contains all logarithms $\ln m_0/m_Q$, while $\Delta\sigma_{\mathrm{R}}(m_Q, m_0)$ is a remainder that is finite for $m_0 \to 0$.

When three partons are produced in the final state (NLO real contribution) the loga-

rithmically enhanced contribution $\Delta\sigma_{\rm L}^{\rm FEX}(m_Q, m_0)$ due to FEX is given by

$$\begin{aligned}\Delta\sigma_{\rm L}^{\rm FEX}(m_Q, m_0) = \frac{\alpha_s}{2\pi}\ln\frac{m_0^2}{m_Q^2} \times \int dx_1 dx_2 \Big[&(P_{Qg}\otimes g)(x_1) g(x_2) \hat{\sigma}^{(0)}_{Qg\to Qg}(x_1,x_2) \\ &+ g(x_1)(P_{Qg}\otimes g)(x_2) \hat{\sigma}^{(0)}_{gQ\to Qg}(x_1,x_2) \\ &+ (P_{Qg}\otimes g)(x_1) q(x_2)\, \hat{\sigma}^{(0)}_{Qq\to Qq}(x_1,x_2) \\ &+ q(x_1)(P_{Qg}\otimes g)(x_2)\, \hat{\sigma}^{(0)}_{qQ\to Qq}(x_1,x_2)\Big]\,, \end{aligned} \tag{A.1}$$

where the first two terms are due to diagrams where the hard scattering process is $Qg \to Qg$, while the remaining terms correspond to diagrams where the hard scattering process is $Qq \to Qq$ and $\hat{\sigma}^{(0)}_{ab\to cd}(x_1, x_2)$ denotes the Born partonic cross section for the process $ab \to cd$ as a function of the incoming energy fractions x_1, x_2. The sums over light-quark flavours (and over quarks and antiquarks) are implicit.

In the case of NLO virtual corrections, for calculations in which the heavy-quark flavour is decoupled both in the running coupling and the PDFs, the only logarithmically enhanced contribution comes from the subprocess $q\bar{q} \to Q\bar{Q}$ (FCR):

$$\Delta\sigma_{\rm L}^{\rm FCR}(m_Q, m_0) \;=\; \frac{2\,\alpha_s T_R}{3\pi}\;\ln\frac{m_0^2}{m_Q^2}\,\sigma^{(0)}_{q\bar{q}\to Q\bar{Q}}\,. \tag{A.2}$$

In the $gg \to Q\bar{Q}$ subprocess, logarithmically enhanced virtual corrections from the renormalisation group evolution of the coupling and the gluon distribution cancel.

As an example, in figure 7 we plot $\sigma(m_Q)$, the integrated inclusive p_t spectrum at the Tevatron for $p_t > 50\,\text{GeV}$ and $|y| < 0.7$, as a function of m_Q for the real and virtual NLO contributions, $\mathcal{O}(\alpha_s^3)$, as given by MCFM. We also show the LO result for reference.

We see that in the small mass region, the cross sections computed with MCFM are well approximated by $\sigma(m_0) + \Delta\sigma_{\rm L}(m_Q, m_0)$, where $\Delta\sigma_{\rm L}(m_Q, m_0)$ is the sum of the finite-mass logarithmic contributions in eq. (A.1) and eq. (A.2) (the integration has been performed numerically with CAESAR [40]). We obtain similar results at the LHC. Finite-mass effects for the inclusive p_t spectra can then be computed by considering the difference $p_t\, d\sigma(m_Q)/dp_t - p_t\, d\sigma(m_0)/dp_t$, where m_0 is as close to zero as numerically possible given the presence of small-mass instabilities in the NLO calculation (we choose $m_0 = 0.2\,\text{GeV}$ at the Tevatron and $m_0 = 1.0\,\text{GeV}$ at the LHC) and subtracting all collinear enhancements predicted from eqs. (A.1) and (A.2). The results of this procedure is what is presented for b-jets in the 4th panel of figure 3. In this manner, we obtained the coefficient $C(p_t)$, as used in eq. (3.1):

$$C(p_t)\ln\frac{m_b}{m_0} = -\frac{d}{dp_t}\Delta\sigma_{\rm L}(m_b, m_0)\,. \tag{A.3}$$

B. Electroweak corrections

There has been discussion in the recent literature [41 – 44] of potentially large electroweak (EW) corrections to QCD light and heavy (top) dijet cross sections. Generally speaking there is consensus that these effects should be modest ($\lesssim 5\%$) at the Tevatron, but it is

JHEP07(2007)026

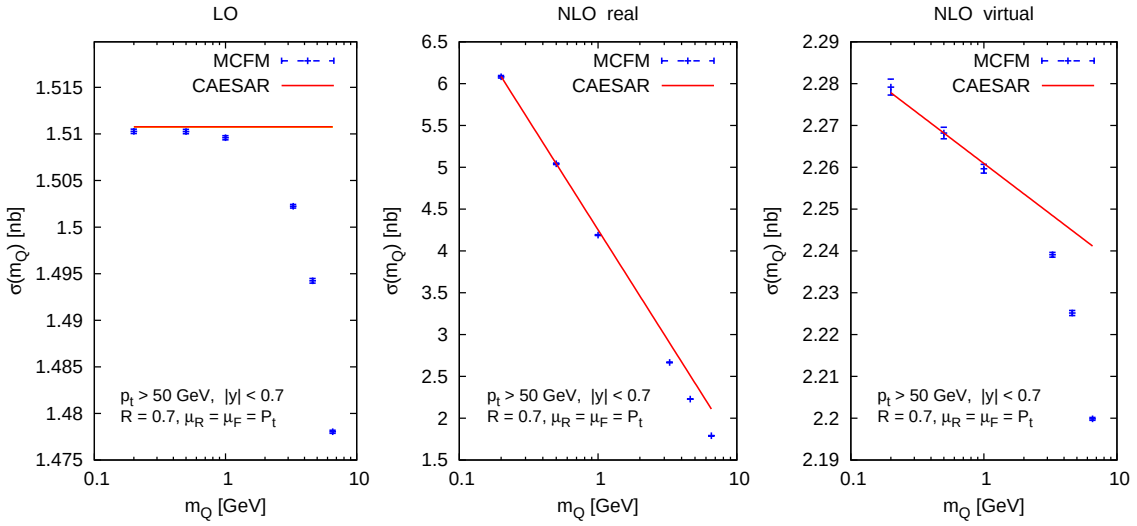

Figure 7: Various contributions to the inclusive cross section for b jets with $p_t > 50\,\text{GeV}$ and $|y| < 0.7$ at the Tevatron, as a function of the heavy-quark mass m_Q. The points are from a massive calculation using MCFM, while at NLO the lines are given by the slopes in eqs. (A.1), (A.2), with a constant term adjusted so as to match the massive calculation at $m_Q = 0.5\,\text{GeV}$.

not uncommon for effects of up to 40% to be quoted at the upper end (4 TeV) of the p_t reach of the LHC.

Two kinds of issues need to be addressed. Firstly there are effects that apply equally to inclusive and flavoured jet cross sections: it has been known for some time now [45 – 47] that electroweak loop corrections for high-p_t processes involve enhancements proportional to $\alpha_{EW}^n \ln^{2n}(p_t/M_W)$. Such terms are analogous to Sudakov double logarithms in QCD, with the difference that the masses at the electroweak scale regulate the infrared and collinear divergences. Because of their double logarithmic structure they become large at high p_t, and they are the main culprits in the 40% effects quoted in [41] at the high end of the LHC reach (4 TeV). A point emphasised there is that a phenomenological understanding of the impact of EW effects also requires that one consider the experimental treatment of real EW radiation. Ref. [42] examined isolated W and Z radiation and found that it compensated for about a quarter of the loop effects. However, the dominant real radiation contribution should come from (soft) collinear W and Z emission, and it is to be expected that this will compensate a significant remaining part of the loop effects.

A second issue arises specifically when considering flavoured jets, because by isolating a given flavour one breaks the electroweak SU(2) symmetry: while the emission of a soft W boson has little effect on the energy of the jet and so should largely cancel with corresponding virtual corrections in the inclusive jet spectrum, if the W is emitted from a b-jet, it will convert it into a top-quark jet [48]. This is often referred to as Bloch-Nordsieck violation [45] and may lead to significant double logarithmic EW corrections for the very highest p_t flavoured jets at the LHC. As for inclusive jet analyses, the details of the experimental treatment are likely to be crucial, since the flavour attributed to the jet will depend on whether the top quark is reconstructed or whether it is only the b-hadron from

JHEP07(2007)026

the $t \to b + W$ decay that is identified. For charm jets the experimental situation will be different insofar as the real EW emission process is $c \to s + W$, and strange hadrons do not decay back to charmed hadrons! The question of flavour-changing EW effects is relevant also for the gluon splitting process, e.g. $g \to c\bar{c}$, where one of the charm quarks may then emit a W, giving a jet with a net charm flavour. If the W is not identifiable experimentally, then at high p_t at LHC this process, which has enhancements of the form $\alpha_s^m \alpha_{EW}^n \ln^{2m-1+2n}(p_t/M_W)$, may give a non-negligible contribution to the charm-jet spectrum.

For both b and c jets, if the experiments prove to be able to measure heavy flavour at these high p_t values, then it will become important to examine the above issues in more detail.

References

[1] ATLAS Physics Technical Design Report, Vol.II, CERN/LHCC 1999-15;
CMS collaboration, *CMS Physics Technical Design Report, Vol. II*, CERN/LHCC 2006-021.

[2] CDF collaboration, A. Abulencia et al., *Measurement of the b jet cross-section in events with a Z boson in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Rev.* **D 74** (2006) 032008 [hep-ex/0605099];
D0 collaboration, Y.D. Mutaf, *Measurement of the ratio of inclusive cross sections sigma($p\bar{p} \to Z + b$)/σ($p\bar{p} \to Z + j$) at D0 run II*, hep-ex/0409039;
D0 collaboration, V.M. Abazov et al., *A search for $Wb\bar{b}$ and WH production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Rev. Lett.* **94** (2005) 091802 [hep-ex/0410062];
CDF colalboration, A. Abulencia et al., *Search for $H \to b\bar{b}$ produced in association with W bosons in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Rev. Lett.* **96** (2006) 081803 [hep-ex/0512051];
J. Campbell, R.K. Ellis, F. Maltoni and S. Willenbrock, *Production of a Z boson and two jets with one heavy-quark tag*, *Phys. Rev.* **D 73** (2006) 054007 [hep-ph/0510362].

[3] CDF collaboration, *Measurement of the inclusive b-jet cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96\,TeV$*, Note 8418;
see also CDF and D0 collaboration, M. D'Onofrio et al., *Jet properties at the Tevatron*, FERMILAB-CONF-06-224-E.

[4] S. Frixione and M.L. Mangano, *Heavy-quark jets in hadronic collisions*, *Nucl. Phys.* **B 483** (1997) 321 [hep-ph/9605270].

[5] CDF - Run II collaboration, A. Abulencia et al., *Measurement of the inclusive jet cross section using the k_T algorithm in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV with the CDF II detector*, *Phys. Rev.* **D 75** (2007) 092006 [hep-ex/0701051].

[6] G.C. Blazey et al., *Run II jet physics*, hep-ex/0005012.

[7] TeV4LHC QCD Working Group collaboration, M.G. Albrow et al., *Tevatron-for-LHC report of the QCD working group*, hep-ph/0610012.

[8] G.P. Salam and G. Soyez, *A practical seedless infrared-safe cone jet algorithm*, arXiv:0704.0292.

[9] S. Catani, Y.L. Dokshitzer, M.H. Seymour and B.R. Webber, *Longitudinally invariant k_t clustering algorithms for hadron hadron collisions*, *Nucl. Phys.* **B 406** (1993) 187;
S.D. Ellis and D.E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev.* **D 48** (1993) 3160 [hep-ph/9305266].

[10] http://mcfm.fnal.gov/;
see also J.M. Campbell and R.K. Ellis, *Radiative corrections to $Zb\bar{b}$ production*, *Phys. Rev.* **D 62** (2000) 114012 [hep-ph/0006304].

[11] G. Marchesini et al., *HERWIG: a Monte Carlo event generator for simulating hadron emission reactions with interfering gluons. Version 5.1 - April 1991*, *Comput. Phys. Commun.* **67** (1992) 465;
G. Corcella et al., *HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, *JHEP* **01** (2001) 010 [hep-ph/0011363].

[12] S. Frixione and B.R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **06** (2002) 029 [hep-ph/0204244];
S. Frixione, P. Nason and B.R. Webber, *Matching NLO QCD and parton showers in heavy flavour production*, *JHEP* **08** (2003) 007 [hep-ph/0305252].

[13] V.N. Gribov and L.N. Lipatov, *Deep inelastic ep scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 438 [*Yad. Fiz.* **15** (1972) 781];
G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, *Nucl. Phys.* **B 126** (1977) 298;
Y.L. Dokshitzer, *Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics. (in russian)*, *Sov. Phys. JETP* **46** (1977) 641 [*Zh. Eksp. Teor. Fiz.* **73** (1977) 1216].

[14] A.H. Mueller and P. Nason, *Heavy particle content in QCD jets*, *Nucl. Phys.* **B 266** (1986) 265;
M.L. Mangano and P. Nason, *Heavy quark multiplicities in gluon jets*, *Phys. Lett.* **B 285** (1992) 160;
M.H. Seymour, *Heavy quark pair multiplicity in e^+e^- events*, *Nucl. Phys.* **B 436** (1995) 163;
Heavy quark production in jets, *Z. Physik* **C 63** (1994) 99.

[15] G. Marchesini and A.H. Mueller, *BFKL dynamics in jet evolution*, *Phys. Lett.* **B 575** (2003) 37 [hep-ph/0308284];
G. Marchesini and E. Onofri, *Exact solution of BFKL equation in jet-physics*, *JHEP* **07** (2004) 031 [hep-ph/0404242].

[16] A. Banfi, G.P. Salam and G. Zanderighi, *Infrared safe definition of jet flavour*, *Eur. Phys. J.* **C 47** (2006) 113 [hep-ph/0601139].

[17] G.P. Salam and D. Wicke, *Hadron masses and power corrections to event shapes*, *JHEP* **05** (2001) 061 [hep-ph/0102343].

[18] J. Campbell, R.K. Ellis, F. Maltoni and S. Willenbrock, *Higgs boson production in association with a single bottom quark*, *Phys. Rev.* **D 67** (2003) 095002 [hep-ph/0204093];
Associated production of a Z boson and a single heavy-quark jet, *Phys. Rev.* **D 69** (2004) 074021 [hep-ph/0312024].

[19] S. Weinzierl, *The forward-backward asymmetry at NNLO revisited*, *Phys. Lett.* **B 644** (2007) 331 [hep-ph/0609021].

JHEP07(2007)026

[20] M.L. Mangano, P. Nason and G. Ridolfi, *Heavy quark correlations in hadron collisions at next-to-leading order*, *Nucl. Phys.* **B 373** (1992) 295.

[21] Z. Nagy, *Three-jet cross sections in hadron hadron collisions at next-to-leading order*, *Phys. Rev. Lett.* **88** (2002) 122003 [hep-ph/0110315]; *Next-to-leading order calculation of three-jet observables in hadron hadron collision*, *Phys. Rev.* **D 68** (2003) 094002 [hep-ph/0307268].

[22] D. Stump et al., *Inclusive jet production, parton distributions and the search for new physics*, *JHEP* **10** (2003) 046 [hep-ph/0303013].

[23] T. Carli, G.P. Salam and F. Siegert, *A posteriori inclusion of PDFS in NLO QCD final-state calculations*, hep-ph/0510324;
T. Carli, D. Clements et al., in preparation.

[24] M. Dasgupta and G.P. Salam, *Resummation of the jet broadening in DIS*, *Eur. Phys. J.* **C 24** (2002) 213 [hep-ph/0110213];
G.P. Salam, *Higher Order Perturbative Parton Evolution Toolkit, (HOPPET)*, http://projects.hepforge.org/hoppet/.

[25] M.R. Whalley, D. Bourilkov and R.C. Group, *The Les Houches accord PDFs (LHAPDF) and LHAGLUE*, hep-ph/0508110.

[26] M. Cacciari, M. Greco and P. Nason, *The p_T spectrum in heavy-flavour hadroproduction*, *JHEP* **05** (1998) 007 [hep-ph/9803400].

[27] A.D. Martin, R.G. Roberts, W.J. Stirling and R.S. Thorne, *Uncertainties of predictions from parton distributions. I: experimental errors*, *Eur. Phys. J.* **C 28** (2003) 455 [hep-ph/0211080].

[28] W.K. Tung et al., *Heavy quark mass effects in deep inelastic scattering and global QCD analysis*, *JHEP* **02** (2007) 053 [hep-ph/0611254].

[29] A.D. Martin, R.G. Roberts, W.J. Stirling and R.S. Thorne, *Physical gluons and high E_T jets*, *Phys. Lett.* **B 604** (2004) 61 [hep-ph/0410230].

[30] J. Pumplin, H.L. Lai and W.K. Tung, *The charm parton content of the nucleon*, *Phys. Rev.* **D 75** (2007) 054029 [hep-ph/0701220].

[31] S.J. Brodsky, P. Hoyer, C. Peterson and N. Sakai, *The intrinsic charm of the proton*, *Phys. Lett.* **B 93** (1980) 451.

[32] F.S. Navarra, M. Nielsen, C.A.A. Nunes and M. Teixeira, *On the intrinsic charm component of the nucleon*, *Phys. Rev.* **D 54** (1996) 842 [hep-ph/9504388];
S. Paiva, M. Nielsen, F.S. Navarra, F.O. Duraes and L.L. Barz, *Virtual meson cloud of the nucleon and intrinsic strangeness and charm*, *Mod. Phys. Lett.* **A 13** (1998) 2715 [hep-ph/9610310].

[33] CDF collaboration, D. Acosta et al., *Measurements of $b\bar{b}$ azimuthal production correlations in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, *Phys. Rev.* **D 71** (2005) 092001 [hep-ex/0412006].

[34] A. Bocci, *Jet flavour tagging with the CMS experiment*, CMS TS-2006/004;
C. Weiser, *A combined secondary vertex based B-tagging algorithm in CMS*, CMS NOTE-2006/014.

[35] ATLAS collaboration, M. Sapinski et al., *Expected performance of ATLAS for measurements of jets, b-jets, tau-jets, and ETmis*, SN-ATLAS-2002-012, ATL-COM-CONF-2001-006, *Eur. Phys. J.* **C 4S1** (2002) 08.

JHEP07(2007)026

[36] J. Bastos, *Tagging heavy flavours with boosted decision trees*, physics/0702041.

[37] CDF and D0 collaborations, T. Kuhl, *Beauty and charm production cross section measurements at the Tevatron*, talk given at Moriond QCD, March 2007, http://moriond.in2p3.fr/QCD/2007/MondayAfternoon/Kuhl.pdf.

[38] H.L. Lai et al., *The strange parton distribution of the nucleon: global analysis and applications*, *JHEP* **04** (2007) 089 [hep-ph/0702268].

[39] S. Dittmaier, P. Uwer and S. Weinzierl, *NLO QCD corrections to $t\bar{t}$ + jet production at hadron colliders*, hep-ph/0703120.

[40] A. Banfi, G.P. Salam and G. Zanderighi, *Principles of general final-state resummation and automated implementation*, *JHEP* **03** (2005) 073 [hep-ph/0407286].

[41] S. Moretti, M.R. Nolten and D.A. Ross, *Weak corrections to four-parton processes*, *Nucl. Phys.* **B 759** (2006) 50 [hep-ph/0606201].

[42] U. Baur, *Weak boson emission in hadron collider processes*, *Phys. Rev.* **D 75** (2007) 013005 [hep-ph/0611241].

[43] J.H. Kuhn, A. Scharf and P. Uwer, *Electroweak effects in top-quark pair production at hadron colliders*, *Eur. Phys. J.* **C 51** (2007) 37 [hep-ph/0610335];
J.H. Kuhn, A. Scharf and P. Uwer, *Electroweak corrections to top-quark pair production in quark-antiquark annihilation*, *Eur. Phys. J.* **C 45** (2006) 139 [hep-ph/0508092].

[44] S. Moretti, M.R. Nolten and D.A. Ross, *Weak corrections to gluon-induced top-antitop hadro-production*, *Phys. Lett.* **B 639** (2006) 513 [hep-ph/0603083].

[45] M. Ciafaloni, P. Ciafaloni and D. Comelli, *Bloch-Nordsieck violating electroweak corrections to inclusive TeV scale hard processes*, *Phys. Rev. Lett.* **84** (2000) 4810 [hep-ph/0001142];
M. Ciafaloni, P. Ciafaloni and D. Comelli, *Electroweak Bloch-Nordsieck violation at the TeV scale: 'strong' weak interactions?*, *Nucl. Phys.* **B 589** (2000) 359 [hep-ph/0004071];
Bloch-Nordsieck violation in spontaneously broken abelian theories, *Phys. Rev. Lett.* **87** (2001) 211802 [hep-ph/0103315].

[46] M. Melles, *Electroweak radiative corrections in high energy processes*, *Phys. Rept.* **375** (2003) 219 [hep-ph/0104232].

[47] A. Denner, *Electroweak radiative corrections at high energies*, hep-ph/0110155.

[48] P. Ciafaloni and D. Comelli, *The importance of weak bosons emission at LHC*, *JHEP* **09** (2006) 055 [hep-ph/0604070].

JHEP07(2007)026