Fermilab FERMILAB-CONF-07-706-T

Some Interesting Min-Bias Distributions for Early LHC Runs

P. Z. Skands[1,2]
[1]CERN PH-TH, CH-1211 Geneva 23, Switzerland
[2]Theoretical Physics, Fermilab MS106, Box 500, Batavia, IL–60510, USA

Abstract
A few observable distributions in min-bias (inelastic, non-diffractive) events which could be well constrained with early LHC data are presented, with some comments on their significance for placing constraints on theoretical models. The effects of fiducial cuts ($p_\perp > 0.5\,\text{GeV}$, $|\eta| < 2.5$) and extrapolation from the Tevatron are illustrated. Additional interesting plots are collected at:
`http://home.fnal.gov/~skands/leshouches-plots/`

1. INTRODUCTION

At first glance, the confined nature of both the initial and final state implies that there are no perturbatively calculable observables in inelastic hadron-hadron collisions. Under ordinary circumstances, however, two powerful tools are used to circumvent this problem, factorisation and infrared safety. The trouble with minimum-bias and underlying-event (MB/UE) physics is that the applicability of both of these tools is, at best, questionable for a wide range of interesting observables.

To understand why the main perturbative tools are ineffective, let us begin with factorisation. When applicable, factorisation allows us to subdivide the calculation of an observable (regardless of whether it is infrared safe or not) into a perturbatively calculable short-distance part and a universal long-distance part, the latter of which may be modeled and constrained by fits to data. However, in the context of hadron collisions the oft made separation into "hard scattering" and "underlying event" components is not necessarily equivalent to a clean separation in terms of formation/fluctuation time, since the underlying event may contain short-distance physics of its own. Regardless of which definition is more correct, any breakdown of the assumed factorisation could introduce a process-dependence of the long-distance part, leading to an unknown systematic uncertainty in the procedure of measuring the corrections in one process and applying them to another.

The second tool, infrared safety, provides us with a class of observables which are insensitive to the details of the long-distance physics. This works up to corrections of order the long-distance scale divided by the short-distance scale, $Q^n_{\rm IR}/Q^n_{\rm UV}$, where the power n depends on the observable in question and $Q_{\rm IR,UV}$ denote generic infrared and ultraviolet scales in the problem. Since $Q_{\rm IR}/Q_{\rm UV} \to 0$ for large $Q_{\rm UV}$, such observables "decouple" from the infrared physics as long as all relevant scales are $\gg Q_{\rm IR}$. Infrared sensitive quantities, on the other hand, contain logarithms $\log^n(Q^2_{\rm UV}/Q^2_{\rm IR})$ which grow increasingly large as $Q_{\rm IR}/Q_{\rm UV} \to 0$. In MB/UE studies, many of the important measured distributions are not infrared safe in the perturbative sense. Take particle multiplicities, for instance; in the absence of non-trivial infrared effects, the number of partons that would be mapped to hadrons in a naïve local-parton-hadron-duality [1] picture depends logarithmically on the infrared cutoff.

We may thus classify collider observables in four categories: least intimidating are the factorisable infrared safe quantities, such as the R ratio in e^+e^- annihilation, which are only problematic at low scales (where the above-mentioned power corrections can be large). Then come the factorisable infrared sensitive quantities, with the long-distance part parametrised by process-independent non-perturbative functions, such as parton distributions. Somewhat nastier are non-factorised infrared safe observables. An example could here be the energy flow into one of Rick Field's "transverse regions" [2]. The energy flow is nominally infrared safe, but in these regions where bremsstrahlung is

suppressed there can be large contributions from pairwise balancing minijets which are correlated to the hard scattering and hence do not factorise according to at least one of the definitions outlined above (see also [3, 4]). The nastiest beasts by all accounts are non-factorised infrared sensitive quantities, such as the particle multiplicity in the transverse region.

The trouble, then, is that MB/UE physics is full of distributions of the very nastiest kinds imaginable. Phenomenologically, the implication is that the theoretical treatment of non-factorised and non-perturbative effects becomes more important and the interpretation of experimental distributions correspondingly more involved. The problem may also be turned around, noting that MB/UE offers an ideal lab for studying these theoretically poorly understood phenomena; the most interesting observables and cuts, then, are those which minimise the "backgrounds" from better-known physics.

As part of the effort to spur more interplay between theorists and experimentalists in this field, we here present a collection of simple min-bias distributions that carry interesting and complementary information about the underlying physics, both perturbative and non-perturbative. The main point is that, while each plot represents a complicated cocktail of physics effects, such that most models could probably be tuned to give an acceptable description observable by observable, it is very difficult to simultaneously describe the entire set. It should therefore be possible to carry out systematic physics studies beyond simple tunings. For brevity, this text only includes a representative selection, with more results available on the web [5]. Note also that we have here left out several important ingredients which are touched on elsewhere in these proceedings, such as observables involving explicit jet reconstruction and observables in leading-jet, dijet, jet + photon, and Drell-Yan events. See also the underlying-event sections in the HERA-and-the-LHC [6] and Tevatron-for-LHC [7] writeups.

2. MODELS

We have chosen to consider a set of six different tunes of the PYTHIA event generator [8], called A, DW, and DWT [2, 7], S0 and S0A [9], and ATLAS-DC2 / Rome [10]. For min-bias, all of these start from leading order QCD $2 \to 2$ matrix elements, augmented by initial- and final-state showers (ISR and FSR, respectively) and perturbative multiple parton interactions (MPI) [11, 12], folded with CTEQ5L parton distributions [13] on the initial-state side and the Lund string fragmentation model [14] on the final-state side. In addition, the initial state is characterised by a transverse mass distribution roughly representing the degree of lumpiness in the proton[1] and by correlated multi-parton densities derived from the standard ones by imposing elementary sum rules such as momentum conservation [11] and flavour conservation [17]. The final state, likewise, is subject to several effects unique to hadronic collisions, such as the treatment of beam remnants (e.g., affecting the flow of baryon number) and colour (re-)connection effects between the MPI final states [9, 11, 18].

Although not perfectly orthogonal in "model space", these tunes are still reasonably complementary on a number of important points, as illustrated in tab. 1. Column by column in tab. 1, these differences are as follows: 1) showers off the MPI are only included in S0(A). 2) the MPI infrared cutoff scale evolves faster with collision energy in tunes A, DW, and S0A than in S0 and DWT. 3) all models except the ATLAS tune have very strong final-state colour correlations. 4) tunes A, DW(T), and ATLAS use Q^2-ordered showers and the old MPI framework, whereas tunes S0(A) use the new interleaved $p_\perp$-ordered model. 5) tunes A and DW(T) have transverse mass distributions which are significantly more peaked than Gaussians, with ATLAS following close behind, and S0(A) having the smoothest distribution. 6) the models were tuned to describe one or more of min-bias (MB), underlying-event (UE), and/or Drell-Yan (DY) data at the Tevatron.

Tunes DW and DWT only differ in the energy extrapolation away from the Tevatron and hence

[1] Note that the impact-parameter dependence is still assumed factorised from the x dependence in these models, $f(x, b) = f(x)g(b)$, where b denotes impact parameter, a simplifying assumption that by no means should be treated as inviolate, see e.g. [4, 15, 16].

Model	Showers off MPI	MPI $p_\perp$ cutoff at $1.96 \to 14\,\text{TeV}$	FS Colour Correlations	Shower Ordering	Proton Lumpiness	Tevatron Constraints
A	No	$2.04 \overset{\text{fast}}{\to} 3.34$	Strong	Q^2	More	MB, UE
DW	No	$1.94 \overset{\text{fast}}{\to} 3.17$	Strong	Q^2	More	MB, UE, DY
DWT	No	$1.94 \overset{\text{slow}}{\to} 2.66$	Strong	Q^2	More	MB, UE, DY
S0	Yes	$1.88 \overset{\text{slow}}{\to} 2.57$	Strong	$p_\perp^2$	Less	MB, DY
S0A	Yes	$1.89 \overset{\text{fast}}{\to} 3.09$	Strong	$p_\perp^2$	Less	MB, DY
ATLAS	No	$2.00 \overset{\text{slow}}{\to} 2.75$	Weak	Q^2	More	UE

Table 1: Brief overview of models. Note that the IR cutoff in these models is not imposed as a step function, but rather as a smooth dampening, see [11, 12]. The labels $\overset{\text{fast}}{\to}$ and $\overset{\text{slow}}{\to}$ refer to the pace of the scaling of the cutoff with collider energy.

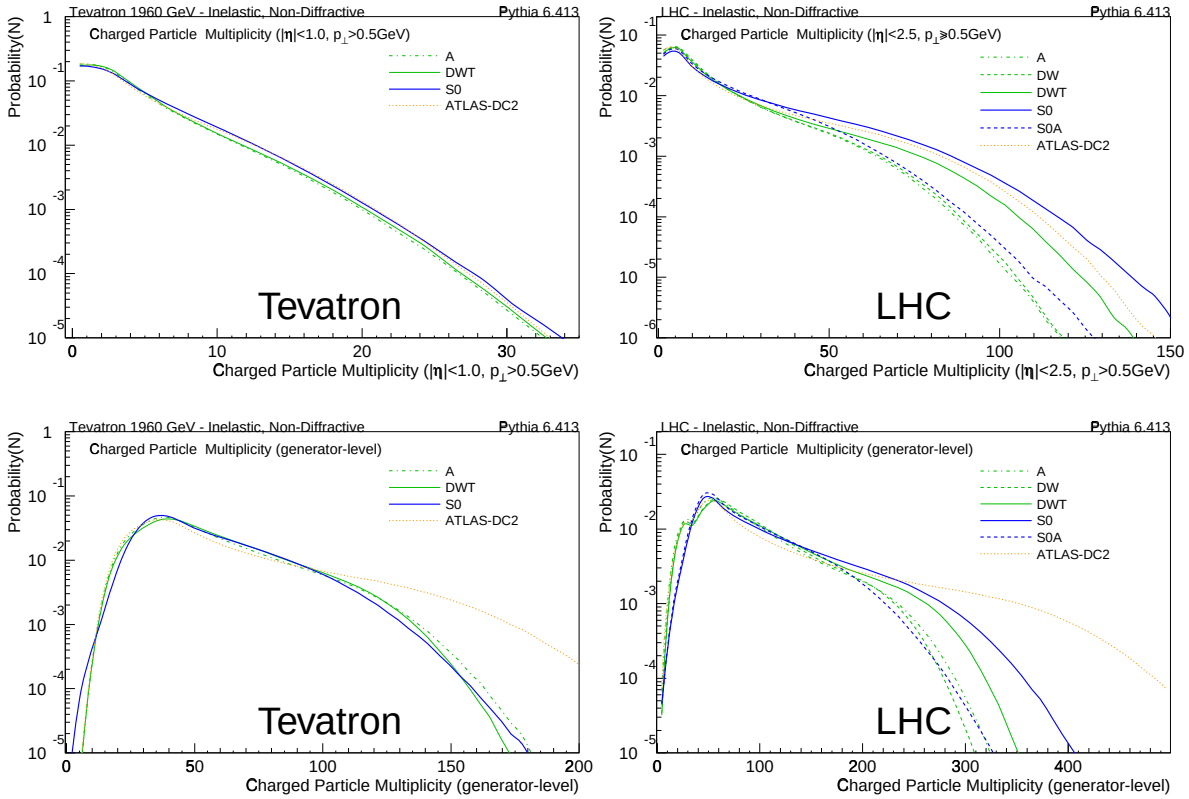

Fig. 1: Charged particle multiplicity distributions, at fiducial (top) and generator (bottom) levels, for the Tevatron (left) and LHC (right). The fiducial averages range from $3.3 < \langle N_{\text{ch}} \rangle < 3.6$ at the Tevatron to $13.0 < \langle N_{\text{ch}} \rangle < 19.3$ at the LHC.

are only shown separately at the LHC. Likewise for S0 and S0A. We regret not including a comparison to other MB/UE Monte Carlo generators, but note that the S0(A) models are very similar to PYTHIA 8 [19], apart from the colour (re-)connection model and some subtleties connected with the parton shower, and that the SHERPA [20] model closely resembles the Q^2-ordered models considered here, with the addition of showers off the MPI. The JIMMY add-on to HERWIG [21, 22] is currently only applicable to underlying-event and not to min-bias.

3. RESULTS

In this section we focus on the following distributions for inelastic non-diffractive events at the Tevatron and LHC: charged particle multiplicity $P(N_{\text{ch}})$, $dN_{\text{ch}}/dp_\perp$, $dN_{\text{ch}}/d\eta$, the average $p_\perp$ vs. N_{ch}

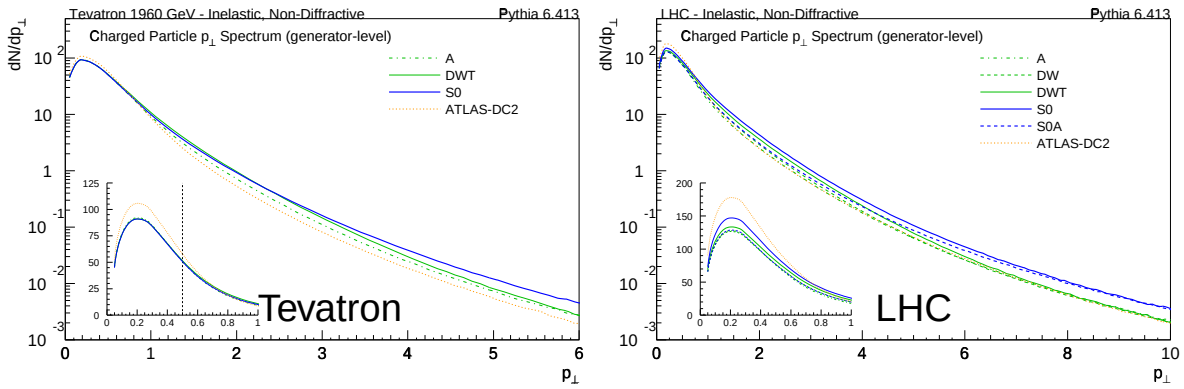

Fig. 2: Charged particle $p_\perp$ spectrum, generator-level only. Insets show the region below 1 GeV on a linear scale. The fiducial distributions [5] are very similar, apart from an overall normalisation and the cut at $p_\perp = 0.5$ GeV.

correlation, the forward-backward $N_{\rm ch}$ and $E_\perp$ correlations vs. η, as well as a few plots of theoretical interest showing the multiplicity distribution of multiple interactions $P(N_{\rm int})$. On most of the plots we include the effects of fiducial cuts, which are represented by the cuts $p_\perp > 0.5\,$GeV and $|\eta| < 1.0$ ($|\eta| < 2.5$) at the Tevatron (LHC).

The charged particle multiplicity is shown in fig. 1, both including fiducial cuts (top row) and at generator-level (bottom row). Tevatron results are shown to the left and LHC ones to the right. Given the amount of tuning that went into all of these models, it is not surprising that there is general agreement on the charged track multiplicity in the fiducial region at the Tevatron (top left plot). In the top right plot, however, it is clear that this near-degeneracy is broken at the LHC, due to the different energy extrapolations, and hence even a small amount of data on the charged track multiplicity will yield important constraints. The bottom row of plots shows how things look at the generator-level, i.e., without fiducial cuts. An important difference between the ATLAS tune and the other models emerges. The ATLAS tune has a significantly higher component of unobserved charged multiplicity. This highlights the fact that extrapolations from the measured distribution to the generator-level one are model-dependent.

The cause for the difference in unobserved multiplicity can be readily identified by considering the generator-level $p_\perp$ spectra of charged particles, fig. 2. The small insets show the region below 1 GeV on a linear scale, with the cut at $p_\perp = 0.5\,$GeV shown as a dashed line. Below the fiducial cut, the ATLAS tune has a significantly larger soft peak than the other models. The S0 model, on the other hand, has a harder distribution in the tail, which also causes S0 to have a slightly larger overall multiplicity in the central region, as illustrated in the fiducial pseudorapidity distributions, fig. 3. Apart from the overall normalisation, however, the pseudorapidity distribution is almost featureless except for the tapering off towards large $|\eta|$ at the LHC. Nonetheless, we note that to study possible non-perturbative fragmentation differences between LEP and hadron colliders, quantities that would be interesting to plot vs. this axis would be strangeness and baryon fractions, such as $N_{K^0_S}/N_{\rm ch}$ and $N_{\Lambda^0}/(N_{\Lambda^0} + N_{\bar{\Lambda}^0})$, as well as the the $p_\perp$ spectra of these particles. With good statistics, also multi-strange baryons would carry interesting information, as has been studied in pp collisions in particular by the STAR experiment [23, 24].

Before going on to correlations, let us briefly consider how the multiplicity is built up in the various models. Fig. 4 shows the probability distribution of the number of multiple interactions. This distribution essentially represents a folding of the multiple-interactions cross section above the infrared cutoff with the assumed transverse matter distribution. Firstly, the ATLAS and Rick Field tunes have almost identical infrared cutoffs and transverse mass profiles and hence look very similar. (Since ATLAS and DWT have the same energy extrapolation, these are the most similar at LHC.)

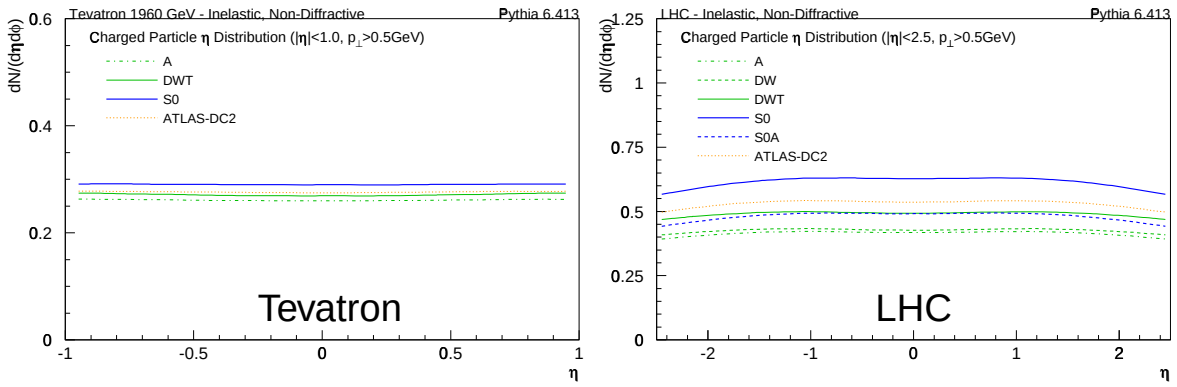

Fig. 3: Charged particle density vs. pseudorapidity, fiducial distribution only. The generator-level ones can be found at [5].

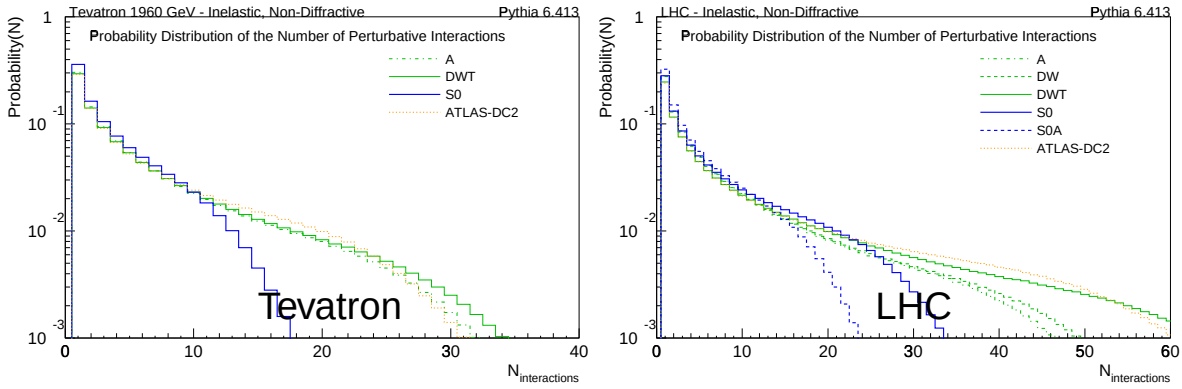

Fig. 4: Probability distribution of the number of multiple interactions. The averages range from $3.7 < \langle N_{\rm int} \rangle < 6.1$ at the Tevatron to $4.7 < \langle N_{\rm int} \rangle < 11.2$ at the LHC.

On the other hand, the S0(A) models exhibit a significantly smaller tail towards large numbers of interactions caused by a combination of the smoother mass profile and the fact that the MPI are associated with ISR showers of their own, hence each takes a bigger x fraction.

Fig. 5 shows the first non-trivial correlation, the average track momentum (counting fiducial tracks only) vs. multiplicity for events with at least one charged particle passing the fiducial cuts. The general trend is that the tracks in high-multiplicity events are harder on average than in low-multiplicity ones. This agrees with collider data and is an interesting observation in itself. We also see that the tunes roughly agree for low-multiplicity events, while the ATLAS tune falls below at high multiplicities. In the models here considered, this is tightly linked to the weak final-state colour correlations in the ATLAS tune; the naive expectation from an uncorrelated system of strings decaying to hadrons would be that $\langle p_\perp \rangle$ should be independent of $N_{\rm ch}$. To make the average $p_\perp$ rise sufficiently to agree with Tevatron data, tunes A, DW(T), and S0(A) incorporate strong colour correlations between final-state partons from different interactions, chosen in such a way as to minimise the resulting string length. An alternative possible explanation could be Cronin-effect-type rescatterings of the outgoing partons, a preliminary study of which is in progress [25].

An additional important correlation, which carries information on local vs. long-distance fluctuations, is the forward-backward correlation strength, b, defined as [11, 26, 27]

$$b = \frac{\langle n_F n_B \rangle - \langle n_F \rangle^2}{\langle n_F^2 \rangle - \langle n_F \rangle^2} \,, \tag{1}$$

where n_F (n_B) is the number of charged particles in a forward (backward) pseudorapidity bin of fixed

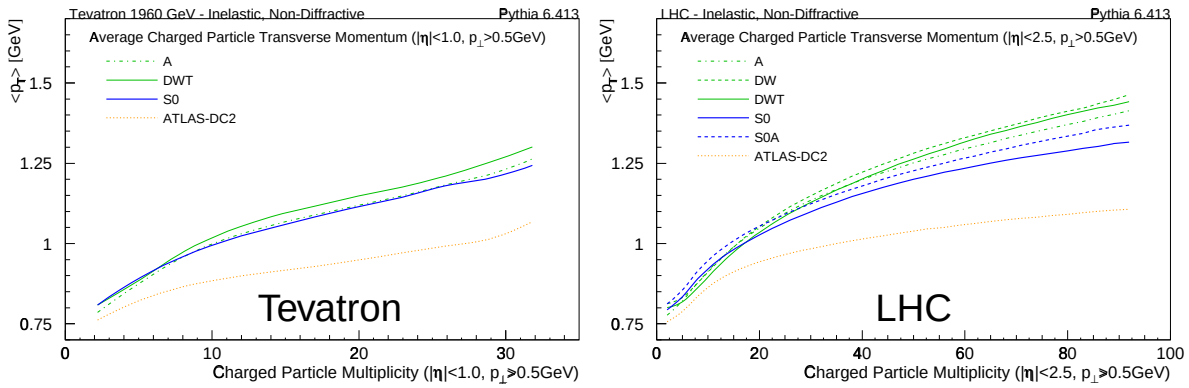

Fig. 5: The average track transverse momentum vs. the number of tracks, counting fiducial tracks only, for events with at least one fiducial track.

size, separated by a central interval $\Delta\eta$ centred at zero. The UA5 study [26] used pseudorapidity bins one unit wide and plotted the correlation vs. the rapidity difference, $\Delta\eta$. For comparison, STAR, which has a much smaller coverage, uses 0.2-unit wide bins [28]. However, as shown in a recent study [29], small bins increase the relative importance of statistical fluctuations, washing out the genuine correlations. For the Tevatron and LHC detectors, which also have small coverages, we therefore settle on a compromise of 0.5-unit wide bins. We also choose to plot the result vs. the pseudorapidity of the forward bin, $\eta_F \sim \Delta\eta/2$, such that the x axis corresponds directly to a pseudorapidity in the detector (the backward bin is then situated symmetrically on the other side of zero). Fig. 6 shows the generator-level correlations, both for charged particles (top row) and for a measure of transverse energy (bottom row), here defined as the $p_\perp$ sum of all neutral and charged particles inside the relevant rapidity bins. Note that we let the x axis extend to pseudorapidities of 5, outside the measurable region, in order to get a more comprehensive view of the behaviour of the distribution. The fact that the ATLAS and S0(A) distributions have a more steeply falling tail than A and DW(T) again reflects the qualitatively different physics cocktails represented by these models. Our tentative conclusions are as follows: Rick Field's tunes A, DW, and DWT have a large number of multiple interactions, cf. fig. 4, but due to the strong final-state colour correlations in these tunes, the main effect of each additional interaction is to add "wrinkles" and energy to already existing string topologies. Their effects on short-distance correlations are therefore suppressed relative to the ATLAS tune, which exhibits similar long-distance correlations but stronger short-distance ones. S0(A) has a smaller total number of MPI, cf. fig. 4, which leads to smaller long-distance correlations, but it still has strong short-distance ones. In summary, the b distributions are clearly sensitive to the relative mix of MPI and shower activity. They also depend on the detailed shape of fig. 4, which in turn is partly controlled by the transverse matter density profile. Measurements of these distributions, both at present and future colliders, would therefore add another highly interesting and complementary piece of information on the physics cocktail.

4. CONCLUSION AND OUTLOOK

We have illustrated some elementary distributions in inelastic, non-diffractive events at the Tevatron and LHC, as they look with various tunes of the two underlying-event models in the PYTHIA event generator. In particular, taking the charged particle multiplicity distribution to set the overall level of the MB/UE physics, the $p_\perp$ spectrum of charged particles and the $\langle p_\perp \rangle (N_{\rm ch})$ correlations then add important information on aspects such as final-state colour correlations. Identified-particle spectra would yield further insight on beam remnants and hadronization in a hadron-collider environment. Finally, correlations in multiplicity and energy vs. pseudorapidity can be used to extract information

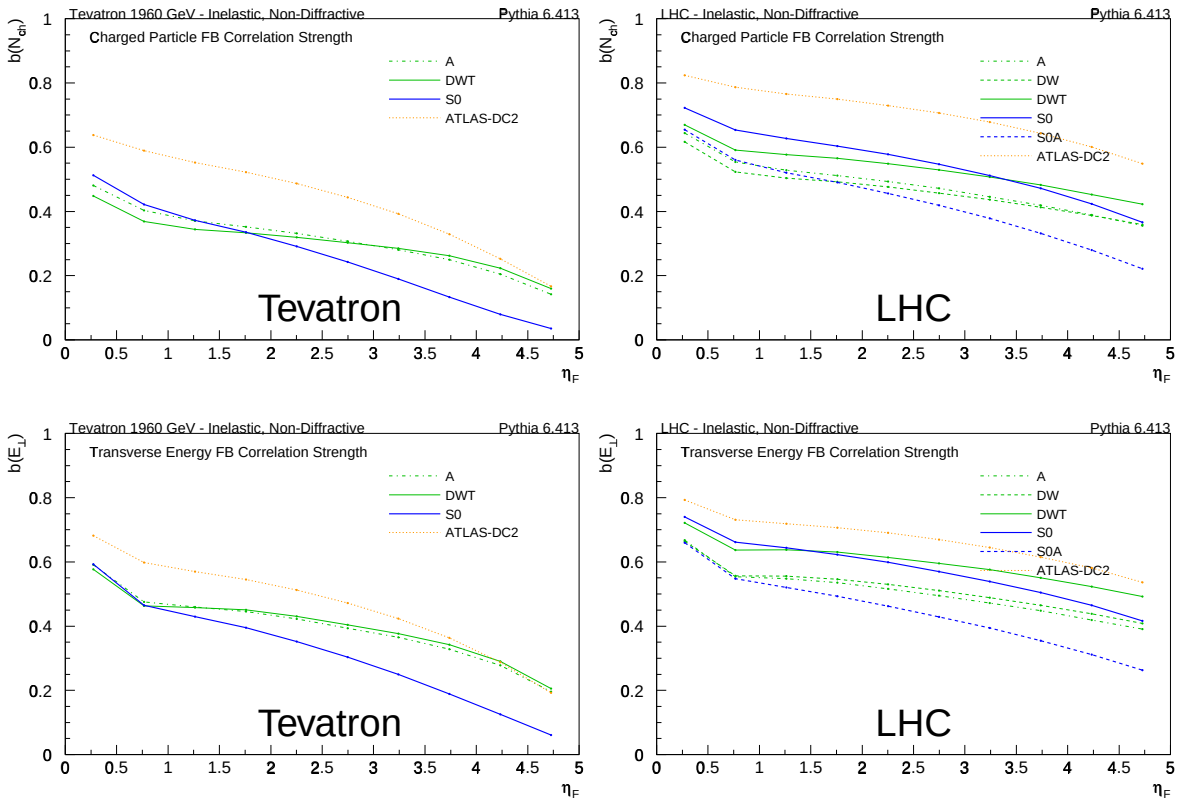

Fig. 6: Generator-level forward-backward correlation strength, b, for charged particles (top) and transverse energy (bottom).

on the importance of short-distance vs. long-distance correlations, which (very) roughly correspond to the type of fluctuations produced by shower- and multiple-interaction-activity, respectively.

By comparing the multiplicity distributions with and without fiducial cuts, we note that the extrapolation from observed to generator-level distributions can be highly model-dependent. It is therefore important to extend the measured region as far as possible in both η and $p_\perp$.

On the phenomenological side, several remaining issues could still be addressed without requiring a more formal footing (see below). These include parton rescattering effects (Cronin effect) [25], correlations between x- and impact-parameter-dependence in the multi-parton PDFs [3, 15, 16], saturation and small-x effects [30], improved modeling of baryon production [17, 31, 32], possible breakdowns of jet universality between LEP, HERA, and hadron colliders, and closer studies of the correspondence between coherent phenomena, such as diffraction and elastic scattering, and inelastic non-diffractive processes [4, 33].

Further progress would seem to require a systematic way of improving on the phenomenological models, both on the perturbative and non-perturbative sides, which necessitates some degree of formal developments in addition to more advanced model building. The correspondence with fixed-order QCD is already being elucidated by parton-shower / matrix-element matching methods, already a well-developed field. Though these methods are currently applied mostly to X+jet-type topologies, there is no reason they should not be brought to bear on MB/UE physics as well. Systematic inclusion of higher-order effects in showers (beyond that offered by "clever choices" of ordering, renormalisation, and kinematic variables) would also provide a more solid foundation for the perturbative side of the calculation, though this is a field still in its infancy [34, 35]. To go further, however, factorisation in the context of hadron collisions needs to be better understood, probably including by now well-established short-distance phenomena such as multiple perturbative interactions on the "short-

distance" side and, correspondingly, correlated multi-parton PDFs on the "long-distance" side. It is also interesting to note that current multiple-interactions models effectively amount to a resummation of scattering cross sections, in much the same way as parton showers represent a resummation of emission cross sections. However, whereas a wealth of higher-order analytical results exist for emission-type corrections, which can be used as useful cross-checks and tuning benchmarks for parton showers, corresponding results for multiple-interactions corrections are almost entirely absent. This is intimately linked to the absence of a satisfactory formulation of factorisation.

On the experimental side, it should be emphasised that there is much more than Monte Carlo tuning to be done in MB/UE studies, and that data is vital to guide us in both the phenomenological and formal directions discussed above. Dedicated Tevatron studies have already had a large impact on our understanding of hadron collisions, but much remains uncertain. Results of future measurements are likely to keep challenging that understanding and could provide for a very fruitful interplay between experiment and theory.

References

[1] Y. I. Azimov, Y. L. Dokshitzer, V. A. Khoze, and S. I. Troyan, *Z. Phys.* **C27** (1985) 65–72.

[2] R. Field and R. C. Group,, **CDF** Collaboration hep-ph/0510198.

[3] V. L. Korotkikh and A. M. Snigirev, *Phys. Lett.* **B594** (2004) 171–176, [hep-ph/0404155].

[4] D. Treleani, *Phys. Rev.* **D76** (2007) 076006, [arXiv:0708.2603 [hep-ph]].

[5] P. Skands,, "Some interesting min-bias distributions for early LHC runs." See http://home.fnal.gov/∼skands/leshouches-plots/.

[6] S. Alekhin *et. al.*, hep-ph/0601012.

[7] M. G. Albrow *et. al.*,, **TeV4LHC QCD Working Group** Collaboration hep-ph/0610012.

[8] T. Sjöstrand, S. Mrenna, and P. Skands, *JHEP* **05** (2006) 026, [hep-ph/0603175].

[9] P. Skands and D. Wicke, *Eur. Phys. J.* **C52** (2007) 133–140, [hep-ph/0703081].

[10] A. Moraes *et. al.*, SN-ATLAS-2006-057.

[11] T. Sjöstrand and M. van Zijl, *Phys. Rev.* **D36** (1987) 2019.

[12] T. Sjöstrand and P. Z. Skands, *Eur. Phys. J.* **C39** (2005) 129–154, [hep-ph/0408302].

[13] H. L. Lai *et. al.*,, **CTEQ** Collaboration *Eur. Phys. J.* **C12** (2000) 375–392, [hep-ph/9903282].

[14] B. Andersson, *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* **7** (1997) 1–471.

[15] D. B. Renner, *J. Phys. Conf. Ser.* **9** (2005) 264–267, [hep-lat/0501005].

[16] P. Hagler *et. al.*,, **LHPC** Collaboration arXiv:0705.4295 [hep-lat].

[17] T. Sjöstrand and P. Z. Skands, *JHEP* **03** (2004) 053, [hep-ph/0402078].

[18] M. Sandhoff and P. Skands,. Presented at Les Houches Workshop on Physics at TeV Colliders, Les Houches, France, 2-20 May 2005, hep-ph/0604120.

[19] T. Sjöstrand, S. Mrenna, and P. Skands, arXiv:0710.3820 [hep-ph].

[20] T. Gleisberg *et. al.*, *JHEP* **02** (2004) 056, [hep-ph/0311263].

[21] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, *Z. Phys.* **C72** (1996) 637–646, [hep-ph/9601371].

[22] G. Corcella *et. al.*, *JHEP* **01** (2001) 010, [hep-ph/0011363].

[23] B. I. Abelev *et. al.*,, **STAR** Collaboration *Phys. Rev.* **C75** (2007) 064901, [nucl-ex/0607033].

[24] M. Heinz,, **STAR** Collaboration arXiv:0707.1510 [hep-ex].

[25] T. Sjöstrand and R. Corke, work in progress.

[26] R. E. Ansorge *et. al.*,, **UA5** Collaboration *Z. Phys.* **C37** (1988) 191–213.

[27] A. Capella, U. Sukhatme, C.-I. Tan, and J. Tran Thanh Van, *Phys. Rept.* **236** (1994) 225–329.

[28] B. K. Srivastava,, **STAR** Collaboration nucl-ex/0702054.

[29] Y.-L. Yan, B.-G. Dong, D.-M. Zhou, X.-M. Li, and B.-H. Sa, arXiv:0710.2187 [nucl-th].

[30] E. Avsar, G. Gustafson, and L. Lönnblad, *JHEP* **01** (2007) 012, [hep-ph/0610157].

[31] T. Sjöstrand and P. Z. Skands, *Nucl. Phys.* **B659** (2003) 243, [hep-ph/0212264].

[32] R. M. Duran Delgado, G. Gustafson, and L. Lönnblad, *Eur. Phys. J.* **C52** (2007) 113–119, [hep-ph/0702241].

[33] G. Gustafson, arXiv:0712.1941 [hep-ph].

[34] K. Kato and T. Munehisa, *Comput. Phys. Commun.* **64** (1991) 67–97.

[35] Z. Nagy and D. E. Soper, *JHEP* **09** (2007) 114, [arXiv:0706.0017 [hep-ph]].