The Underlying Event

C.M.Buttar[1], *J.M.Butterworth*[2], *R.D.Field*[3], *C.Group*[3], *G.Gustafson*[4], *S.Hoeche*[5], *F.Krauss*[5], *A.Moraes*[1], *M.H.Seymour*[6,7], *A.Schalicke*[8], *P.Szczypka*[9], *T.Sjöstrand*[4,7]
[1]Dept. of Physics and Astronomy, University of Glasgow, UK
[2]Dept. of Physics and Astronomy, University College London, UK
[3]Dept. of Physics, University of Florida, USA
[4]Dept. of Theoretical Physics, Lund University, Sweden
[5]Institute for Theoretical Physics, Dresden University of Technology, FRG
[6]School of Physics and Astronomy, University of Manchester, UK
[7]CERN, Switzerland
[8]DESY Zeuthen, FRG
[9]Dept. of Physics, University of Bristol, UK

Abstract
The contributions to working group II: "Multi-jet final states and energy flows" on the underlying event are summarized. The study of the underlying event in hadronic collisions is presented and Monte Carlo tunings based on this are described. New theoretical and Monte Carlo methods for describing the underlying event are also discussed.

1 Introduction

The underlying event is an important element of the hadronic environment within which all physics at the LHC, from Higgs searches to physics beyond the standard model, will take place. Many aspects of the underlying event will be constrained by LHC data when they arrive. However, the physics is so complex, spanning non-perturbative and perturbative QCD and including sensitivities to multi-scale and very low-x physics, that even after LHC switch-on many uncertainties will remain. For this reason, and also for planning purposes, it is critical to have to hand sensible models containing our best physical knowledge and intuition, tuned to all relevant available data.

In this summary of several contributions to the workshop, we first outline the available models in Section 2, most of which are in use at HERA and/or the Tevatron. Recent improvements, some of which were made during the workshop, are also discussed.

Next, current work on tuning these to data is discussed. The underlying event has been extensively studied by CDF and the latest results are presented in Section 3 and compared to predictions from the PYTHIA and HERWIG+JIMMY Monte Carlo generators. The CDF tunings are compared to other tunings based on CDF data and minimum bias data and used to predict the level of underlying events at the LHC in Sections 4 and 5. These reports are very much a snapshot of ongoing work, which will be continued in the follow-up meetings of this workshop and the TeV4LHC workshop.

One major issue in extrapolating the underlying event (UE) to LHC energies is the possible energy dependence of the transverse momentum cut-off between hard and soft scatters, $\hat{p}_T^{\min}$. The need for such a cut-off may be avoided by using the $k_\perp$ factorization scheme as discussed in Section 6, where soft emissions do not contribute to the total cross-section or to the parton density functions (PDFs), but do contribute to the properties of the event. The cross-section for a chain of partonic emission can be extracted from HERA data and can be used to predict the minijet rate or multiple interaction rate in pp or $\mathrm{p\bar{p}}$ collisions. The running of α_s still introduces a cut-off scale between soft and hard chains; however it has been shown that the total cross-section is insensitive to this cut-off and predictions for the mini-jet rate at the LHC are stable. The hadron multiplicity observed in the CDF underlying event data indicates that the string connections in the underlying event are made to minimise the string length. This is the

opposite to what is observed in e^+e^- collisions. The implications for this on the AGK cutting rules is discussed further in Section 6.

This summary ends with a section on conclusions and suggestions for future work.

2 Underlying event models

Several underlying event models are available, at varying stages of development and use. In this section we review the status of thosed discussed during the workshop.

2.1 Multiple Interactions in PYTHIA

The basic implementation of multiple interactions in PYTHIA is almost 20 years old, and many of the key aspects have been confirmed by comparisons with data. In recent years the model has been gradually improved, with junction-string topologies, with flavour-correlated multiparton densities, and with transverse-momentum-ordered showers interleaved with the multiple interactions. However, the "correct" description of colour flow still remains to be found.

The traditional PYTHIA [1,2] model for multiple interactions (MI) [3] is based on a few principles:

1. The naive perturbative QCD $2 \to 2$ cross section is divergent like $\mathrm{d}p_\perp^2/p_\perp^4$ for transverse momenta $p_\perp \to 0$. Colour screening, from the fact that the incoming coloured partons are confined in colour singlet states, should introduce a dampening of this divergence, e.g. by a factor $p_\perp^4/(p_{\perp 0}^2 + p_\perp^2)^2$, where $p_{\perp 0}$ is a free parameter, which comes out to be of the order of 2 GeV.
2. From the thus regularized integrated interaction rate $\sigma_{\mathrm{int}}(E_{\mathrm{cm}}, p_{\perp 0})$ and the nondiffractive cross section $\sigma_{\mathrm{nd}}(E_{\mathrm{cm}})$, the average number of interactions per event can be derived as $\langle n_{\mathrm{int}} \rangle = \sigma_{\mathrm{int}}/\sigma_{\mathrm{nd}}$. With no impact-parameter dependence, the actual number of interactions is given by a Poissonian with mean as above (modulo some corrections coming from $n_{\mathrm{int}} = 0$).
3. More realistically, since hadrons are extended objects, there should be more (average) activity in central collisions than in peripheral ones. By introducing a matter distribution inside a hadron, the overlap between the two incoming hadrons can be calculated as a function of impact parameter b. The number of interactions is now a Poissonian for each b separately, with a mean proportional to the overlap. All events are required to contain at least one interaction; thereby the cross section is automatically dampened for large b. Empirically, the required hadronic impact parameter profile is more peaked at small b than in a Gaussian distribution.
4. It is natural to consider the interactions in an event in order of decreasing $p_\perp$ values. Such a $p_\perp$ ordering has a natural interpretation in terms of formation-time arguments. The generation procedure can conveniently be written in a language similar to that used for parton showers, with the equivalent of a Sudakov form factor being used to pick the next smaller $p_\perp$, given the previous ones. It allows the hardest interaction to be described in terms of conventional PDFs, whereas subsequent ones have to be based on modified PDFs, at the very least reduced by energy–momentum conservation effects. This also reduces the tail of events with very many interactions.
5. Technical limitations lead to several simplifications, such that only the hardest interaction was allowed to develop initial- and final state interactions, and have flavours selected completely freely.
6. Colour correlations between different scatterings cannot be predicted by perturbation theory, but have a direct consequence on the structure of events. One of the most senstive quantities is $\langle p_\perp \rangle (n_{\mathrm{charged}})$. Data here suggest a very strong colour correlation, where the total string length is essentially minimized in the final state.

For a long period of time, only one significant change was made to this scenario:

7. Originally the $p_{\perp 0}$ parameter had been assumed energy-independent. In the wake of the HERA data [4], which led to newer PDF parametrizations having a steeper small-x behaviour than previously assumed, it became necessary to let $p_{\perp 0}$ increase with energy to avoid too steep a rise of the multiplicity. Such an energy dependence can be motivated by colour screening effects [5]. A functional form $p_{\perp 0} \propto s^{\epsilon}$ with $\epsilon \sim 0.08$ is suggested by Pomeron arguments.

Several studies have been presented based on this framework. Some of the recent tuning activities are described elsewhere in this report. The PYTHIA Tune A [6] is a standard reference for much of the current Tevatron underlying-event and minimum-bias physics studies.

In recent years, an effort has been made to go beyond the framework outlined above. Several new or improved components have been introduced.

1. The fragmentation of junction-string topologies has been implemented [7] . Such topologies must be considered when at least two valence quarks are kicked out of an incoming proton beam particle. Here a proton is modelled as a Y-shaped topology, where each valence quarks sits at the end of one of the three legs going out from the middle, the junction. When some ends of this Y are kicked out, also the junction is set in motion. The junction carries no energy or momentum of its own, but it is around the junction that the baryon inheriting the original baryon number will be formed. The junction rest frame is defined by having 120° between the three jets. A number of technical problems have to be overcome in realistic situations, where also gluons may be colour-connected on the three legs, thus giving more complicated space–time evolution patterns.
2. PDFs are more carefully modelled, to take into account the flavour structure of previous interactions [8], not only the overall energy–momentum constraints. Whenever a valence quark is kicked out, the remaining valence PDF of this flavour is rescaled to the new remaining number. When a sea quark is kicked out, an extra "companion" antiquark distribution contribution is inserted, thereby increasing the likelihood that also the antiquark is kicked out.
3. Also remnant flavours are more carefully considered, along with issues such as primordial $k_{\perp}$ values and remnant longitudinal momentum sharing.
4. A few further impact-parameter possibilities are introduced.
5. New transverse-momentum-ordered showers are introduced, both for initial- and final-state radiation (ISR and FSR) [9]. On the one hand, this appears to give an improved description of (hard) multijet production. On the other hand, it allows all evolution to be viewed in terms of a common "time" ordering given by decreasing $p_{\perp}$ values. This is especially critical for the description of MI and ISR, which are in direct competition, in the sense that both mechanisms take momentum out of the incoming beams and thereby require a rescaling of PDF's at later "times". This approach, with interleaved MI and ISR, is illustrated in Fig. 1.

Currently we still make use of two simplifications to the new $p_{\perp}$-ordered framework: (a) the inclusion of FSR is deferred until the MI and ISR have been considered in full, and (b) there is no intertwining, in which two seemingly separate higher-virtuality parton chains turns out to have a common origin when studied at lower $p_{\perp}$ scales. Fortunately there are good reasons why neither of those omitted aspects should be so important.

There is one big remaining unsolved issue in this model, however, namely that of colour flow. If colours are only connected via the fact that the incoming beam remnants are singlets, the correct $\langle p_{\perp} \rangle (n_{\mathrm{charged}})$ behaviour cannot be reproduced whatever variation is tried. It appears necessary to assume that some final-state colour reconnection mechanism tends to reduce the total string length almost to the minimal possible, as was required for Tune A. The most physically reasonable approach, that is yet not too time-consuming to implement, remains to be found. It is possible that also diffractive topologies will need to become a part of this game.

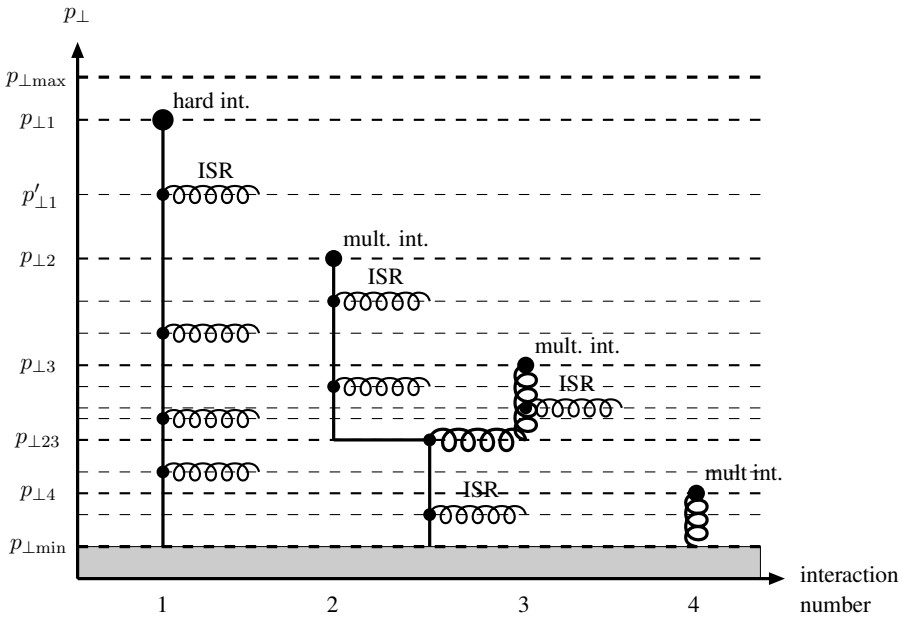


Fig. 1: Schematic figure illustrating one incoming hadron in an event with a hard interaction occurring at $p_{\perp 1}$ and three further interactions at successively lower $p_\perp$ scales, each associated with (the potentiality of) initial-state radiation, and further with the possibility of two interacting partons (2 and 3 here) having a common ancestor in the parton showers. Full lines represent quarks and spirals gluons. The vertical $p_\perp$ scale is chosen for clarity rather than realism; most of the activity is concentrated to small $p_\perp$ values.

Apart from this big colour issue, and the smaller ones of a complete interleaving/intertwining, PYTHIA now contains a very consistent and complete picture of both minimum-bias and underlying-event physics. It will be interesting to see how this framework fares in comparisons with data. However, if the models appears complex, this complexity is driven by necessity: all of the issues already brought up must be included in the "definitive" description, in one form or other, plus possibly some more not yet brought to light.

2.2 JIMMY

The basic ideas of the eikonal model implemented in JIMMY are discussed elsewhere [10]. The model derives from the observation that for partonic scatters above some minimum transverse momentum, $\hat{p}_T^{\min}$, the values of the hadronic momentum fraction x which are probed decrease as the centre-of-mass energy, s, increases, and since the proton structure function rises rapidly at small x [4], high parton densities are probed. Thus the perturbatively-calculated cross section grows rapidly with s. However, at such high densities, the probability of more than one partonic scattering in a single hadron-hadron event may become significant. Allowing such multiple scatters reduces the total cross section, and increases the activity in the final state of the collisions.

2.2.1 Model Assumptions

The JIMMY model assumes some distribution of the matter inside the hadron in impact parameter (b) space, which is independent of the momentum fraction, x. The multiparton interaction rate is then calculated using the cross section for the hard subprocess, the conventional parton densities, and the area overlap function, $A(b)$. No assumption about the behaviour of the *total* cross section is used. For cross sections other than QCD $2 \to 2$ scatters, JIMMY makes use of approximate formulae, valid when all

cross sections except QCD $2 \to 2$ are small, which is true in most cases of interest. This approximation is described in detail elsewhere [11].

2.2.2 *Standard* JIMMY

The starting point for the multiple scattering model is the assertion that, at fixed impact parameter, b, different scatters are independent and so obey Poisson statistics. It is then straightforward to show that the cross section for events in which there are n scatters of type a is given by

$$\sigma_n = \int \mathrm{d}^2 b \, \frac{(A(b)\sigma_{\mathrm{a}})^n}{n!} \, \mathrm{e}^{-A(b)\sigma_{\mathrm{a}}}, \tag{1}$$

where σ_{a} is the parton–parton cross section and $A(b)$ is the matter density distribution, obeying

$$\int \mathrm{d}^2 b \, A(b) = 1. \tag{2}$$

It is straightforward to show that the inclusive cross section for scatters of type a is σ_{a} and the total cross section for events with at least one scatter of type a is

$$\sigma_{\mathrm{tota}} = \int \mathrm{d}^2 b \left(1 - \mathrm{e}^{-A(b)\sigma_{\mathrm{a}}}\right). \tag{3}$$

These can then be combined to give the probability that an event has exactly n scatters of type a, given that it has at least 1 scatter of type a,

$$P_n = \frac{\int \mathrm{d}^2 b \, \frac{(A(b)\sigma_{\mathrm{a}})^n}{n!} \, \mathrm{e}^{-A(b)\sigma_{\mathrm{a}}}}{\int \mathrm{d}^2 b \left(1 - \mathrm{e}^{-A(b)\sigma_{\mathrm{a}}}\right)}, \qquad n \geq 1. \tag{4}$$

This is the probability distribution pretabulated (as a function of $\sqrt{s}$) by Jimmy.

Jimmy's procedure can then be summarized as:

1. Give all events cross section σ_{tota}.
2. In a given event, choose n according to Eq. (4).

It is interesting to note that Jimmy's procedure, despite integrating over b once-and-for-all at initialization time, correctly reproduces the correlation between different scatters, whose physical origin is a b-space correlation: small cross section scatters are more likely to come from events with a large overlap and hence be accompanied by a larger-than-average number of large cross section scatters.

2.2.3 *Two Different Scattering Types*

We consider the possibility that there are two different scattering types, but that the cross section for the second type, σ_{b}, is small enough that events with more than one scatter of type b are negligible. The probability distribution for number of scatters of type a, n, in events with at least one of type b is given by [11]

$$P(n|m \geq 1) = \frac{\int \mathrm{d}^2 b \, \frac{(A(b)\sigma_{\mathrm{a}})^n}{n!} \, \mathrm{e}^{-A(b)\sigma_{\mathrm{a}}} \left(1 - \mathrm{e}^{-A(b)\sigma_{\mathrm{b}}}\right)}{\int \mathrm{d}^2 b \left(1 - \mathrm{e}^{-A(b)\sigma_{\mathrm{b}}}\right)}, \qquad n \geq 0. \tag{5}$$

Since σ_{b} is small, we can expand the exponentials and obtain

$$P(n|m \geq 1) \approx \int \mathrm{d}^2 b \, A(b) \, \frac{(A(b)\sigma_{\mathrm{a}})^n}{n!} \, \mathrm{e}^{-A(b)\sigma_{\mathrm{a}}}, \qquad n \geq 0. \tag{6}$$

Note that this expression is independent of $\sigma_{\rm b}$. It is therefore ideal for implementing into JIMMY. It is useful to rewrite this equation, as follows. We redefine n to be the total number of scatters, including the one of type b (i.e. "new n"="old n"+1) and rewrite, to obtain

$$P_n \approx \frac{\int {\rm d}^2b\, n\, \frac{(A(b)\sigma_{\rm a})^n}{n!}\, {\rm e}^{-A(b)\sigma_{\rm a}}}{\sigma_{\rm a}}\,, \qquad n \geq 1. \tag{7}$$

Note the similarity with Eq. (4), making this form even easier to implement into Jimmy.

The Monte Carlo implementation of this procedure is straightforward:

1. Give all events cross section $\sigma_{\rm b}$.
2. In a given event choose n according to Eq. (7).
3. Generate 1 scatter of type b and $n-1$ of type a.

There is one important difference between the cases in which b is distinct from a and b is a subset of a: some of the $n-1$ scatters of type a could also be of type b. Although this is a small fraction of the total, it can be phenomenologically important. As each scatter of type a is generated, a check is made as to whether it is also of type b. The mth scatter of type b generated so far is rejected with probability $1/(m+1)$. This ensures that the proposed algorithm is continuous at the boundary of b.

When using JIMMY at the LHC, the tuneable parameters are those described previously [10], with the obvious exception of those parameters which only concern the photon. Those remaining are therefore the minimum transverse momentum of a hard scatter, the proton structure, and the effective radius of the proton. Details on how to adjust these parameters can be found elsewhere [11].

2.3 Simulation of Multiple Interactions in `Sherpa`

Given the studies presented in the following sections, and references therein, current multi-purpose event generators rely heavily on the implementation of multiple parton interaction models to describe the final state in hadronic collisions. To allow `Sherpa` to provide a complete description of hadronic events, the module `AMISIC++` has been developed to simulate multiple parton interactions. This module is capable of simulating multiple scatterings according to the formalism initially presented in [3] and in its current implementation acts as a benchmarking tool to cross-check new multiple interaction models [12].

The basic assumption of the multiple interaction formalism according to T. Sjöstrand and M. van Zijl is, that the differential probability $\mathcal{P}(p_\perp^{\rm out})$ to get a (semi-)hard scattering in the underlying event is given by $\mathcal{P}(p_\perp^{\rm out}) = \sigma_{\rm hard}(p_\perp^{\rm out})/\sigma_{\rm ND}$, where $p_\perp^{\rm out}$ is the transverse momentum of the outgoing partons in the scattering. Since $\sigma_{\rm hard}$ is dominated by $2 \to 2$ processes, the definition of $p_\perp^{\rm out}$ is unambiguous. The specific feature of `AMISIC++` is, that it allows for an independent Q^2-evolution of initial and final state partons in each (semi-)hard scattering via an interface to `Sherpa`'s parton shower module `APACIC++` [13, 14]. The key point here is, that the parton shower must then respect the initial $p_\perp^{\rm out}$ distribution of each (semi-)hard scattering. In particular, it must not radiate partons with $p_\perp > p_\perp^{\rm out}$. The appropriate way to incorporate this constraint is in fact identical to the realisation of the highest multiplicity treatment in the CKKW approach [15–18]. Our proposed algorithm works as follows:

1. Create a hard scattering process according to the CKKW approach.
   Employ a K_T jet finding algorithm in the E-scheme to define final state jets.
   Stop the jet clustering as soon as there remains only one QCD node to be clustered.
   Set the starting scale of the multiple interaction evolution to $p_\perp$ of this node.
2. Select $p_\perp$ of the next (semi-)hard interaction according to [3].
   If done for the first time in the event, select the impact parameter b of the collision.

3. Set the jet veto scale of the parton shower to the transverse momentum $p_\perp$, selected in 2. Start the parton shower at the QCD hard scale $\mu^2_{\rm QCD} = 2\,stu/\left(s^2+t^2+u^2\right)$.
4. Return to step 2.

The above algorithm works for pure QCD hard matrix elements as well as for electroweak processes in the hard scattering. In the QCD case the selected starting scale for the determination of the first additional interaction reduces to $p_\perp^{\rm out}$ and is thus equal to the original ordering parameter. In the case of electroweak core processes, like single W- or Z-boson production there is no such unique identification. On the other hand the multiple scatterings in the underlying event must not spoil jet topologies described by the hard event through, e.g., using multi-jet matrix elements. However, since the electroweak bosons may be regarded to have been radiated off QCD partons during the parton shower evolution of a hard QCD event, it is appropriate to reinterprete the hard matrix element as such a QCD+EW process, where the simplest is a 1-jet process.

An important question in conjunction with the simulation of underlying events is the assignment of colours to final state particles. In the `Sherpa` framework, colour connections in any hard $2 \to 2$ QCD process are chosen according to the kinematics of the process. In particular the most probable colour configuration is selected. Additionally, initial state hadrons are considered to be composed from QCD partons in such a way that the colour string lengths in the final state are minimized. In cases, where it is impossible to realise this constraint, the colour configurations of the hard matrix elements are kept but the configuration of the beam remnants is shuffled until a suitable solution is found.

Figures 2–5 show some preliminary results obtained with the above algorithm, implemented in the current `Sherpa` version, `Sherpa-1.0.6`. We compare the `Sherpa` prediction including multiple interactions to the one without multiple interactions and to the result obtained with PYTHIA 6.214, also including multiple interactions and employing the parameters of PYTHIA Tune A [6]. Shown are hadron-level predictions, which are uncorrected for detector acceptance, except for a uniform track finding efficiency as given in [19]. Data were taken at the Fermilab Tevatron during Run I [20]. Good agreement between the simulations and data is observed only if multiple interactions are included. The mean interaction number in `Sherpa`, including the hard scattering, in this case is $<N_{\rm hard}> = 2.08$, while for `PYTHIA 6.214` it is $<N_{\rm hard}> = 7.35$. The lower interaction number in `Sherpa` can easily be understood, as a decrease of parton multiplicity in the (semi-)hard scatterings due to a rise of the parton multiplicity in the parton showers. PYTHIA 6.214 does not allow for parton showers in the (semi-)hard scatterings in the underlying event. This feature has, however, been added in PYTHIA 6.3 (see Section 2.1), and is also present in JIMMY(Section 2.2).

2.4 PHOJET

The physics model used in the MC event generator PHOJET combines the ideas of the DPM [21] with perturbative QCD to give an almost complete picture of high-energy hadron collisions [22].

PHOJET is formulated as a two-component model containing contributions from both soft and hard interactions. The DPM is used to describe the dominant soft processes and perturbative QCD is applied to generate hard interactions.

There has been very little development on PHOJET for the last few years, although it is used quite widely in minimum bias and cosmic ray physics. A major disadvantage for the LHC is that it is not part of a general purpose generator, and therefore cannot be used to generate underlying events to low cross section processes.

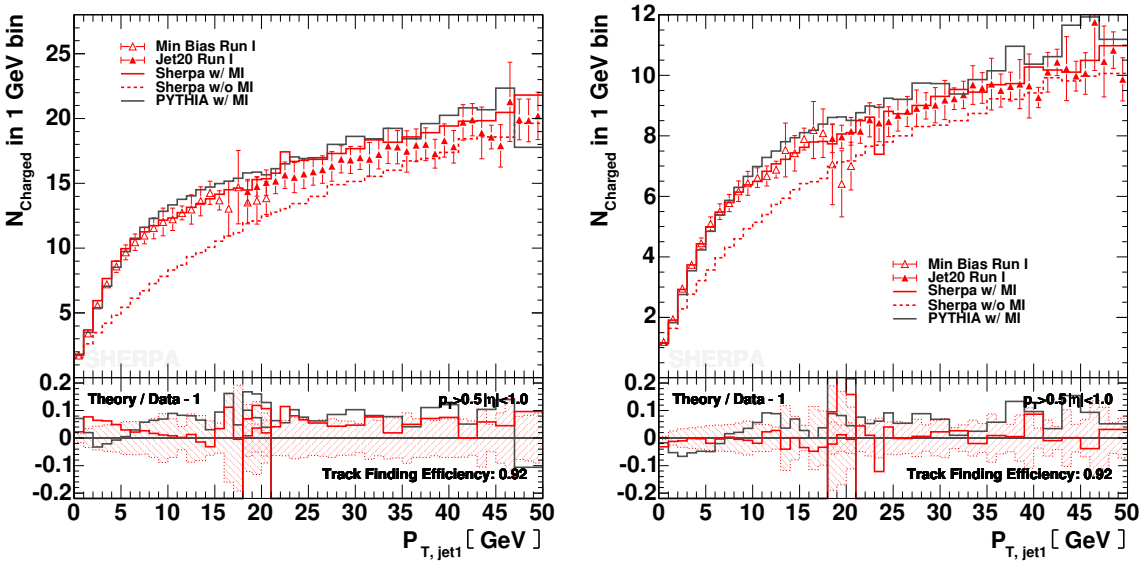


Fig. 2: Charged particle multiplicity as a function of P_T of the leading charged particle jet. The left figure shows the total charged particle multiplicity in the selected p_T- and η-range, the right one displays the same in the "Toward" region (for definitions, see Section 3 and [20]).

3 Tuning PYTHIA and HERWIG/JIMMY in Run 2 at CDF

The behaviour of the charged particle ($p_T > 0.5\,\mathrm{GeV/c}$, $|\eta| < 1$) and energy ($|\eta| < 1$) components of the UE in hard scattering proton-antiproton collisions at $1.96\,\mathrm{TeV}$ has been studied at CDF. The goal is to produce data on the UE that is corrected to the particle level, so that it can be used to tune the QCD Monte-Carlo models using tools such as those described in the contributions from Group 5 of this workshop without requiring a simulation of the CDF detector. Unlike the previous CDF Run 2 UE analysis which used JetClu to define "jets" and compared uncorrected data with the QCD Monte-Carlo models after detector simulation (i.e., CDFSIM), this analysis uses the midpoint jet algorithm and corrects the observables to the particle level. The corrected observables are then compared with the QCD Monte-Carlo models at the particle level (i.e., generator level). The QCD Monte-Carlo models include PYTHIA Tune A, HERWIG and a tuned version of JIMMY.

One can use the topological structure of hadron-hadron collisions to study the UE [19,23,24]. The direction of the leading calorimeter jet is used to isolate regions of η-ϕ space that are sensitive to the UE. As illustrated in Fig. 6, the direction of the leading jet, jet#1, is used to define correlations in the azimuthal angle, $\Delta\phi$. The angle $\Delta\phi = \phi - \phi_{\mathrm{jet\#1}}$ is the relative azimuthal angle between a charged particle (or a calorimeter tower) and the direction of jet#1. The "transverse" region is perpendicular to the plane of the hard 2-to-2 scattering and is therefore very sensitive to the UE. We restrict ourselves to charged particles in the range $p_T > 0.5\,\mathrm{GeV/c}$ and $|\eta| < 1$ and calorimeter towers with $E_T > 0.1\,\mathrm{GeV}$ and $|\eta| < 1$, but allow the leading jet that is used to define the "transverse" region to have $|\eta(jet\#1)| < 2$. Furthermore, we consider two classes of events. We refer to events in which there are no restrictions placed on the second and third highest P_T jets (jet#2 and jet#3) as "leading jet" events. Events with at least two jets with $P_T > 15\,\mathrm{GeV/c}$ where the leading two jets are nearly "back-to-back" ($|\Delta\phi| > 150°$) with $P_T(jet\#2)/P_T(jet\#1) > 0.8$ and $P_T(jet\#3) < 15\,\mathrm{GeV/c}$ are referred to as "back-to-back" events. "Back-to-back" events are a subset of the "leading jet" events. The idea is to suppress hard initial and final-state radiation thus increasing the sensitivity of the "transverse" region to the "beam-beam remnants" and the multiple parton scattering component of the "underlying event".

As illustrated in Fig. 7, we define a variety of MAX and MIN "transverse" regions which help to

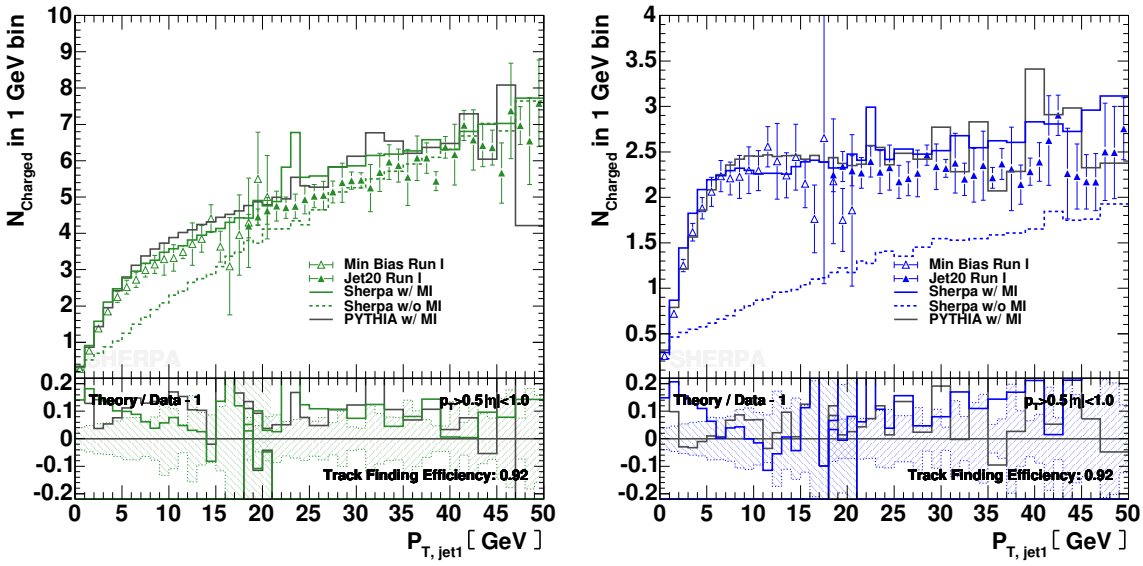


Fig. 3: Charged particle multiplicity as a function of P_T of the leading charged particle jet. The left figure shows results for the "Away" side region, the right one displays results for the "Transverse" region.

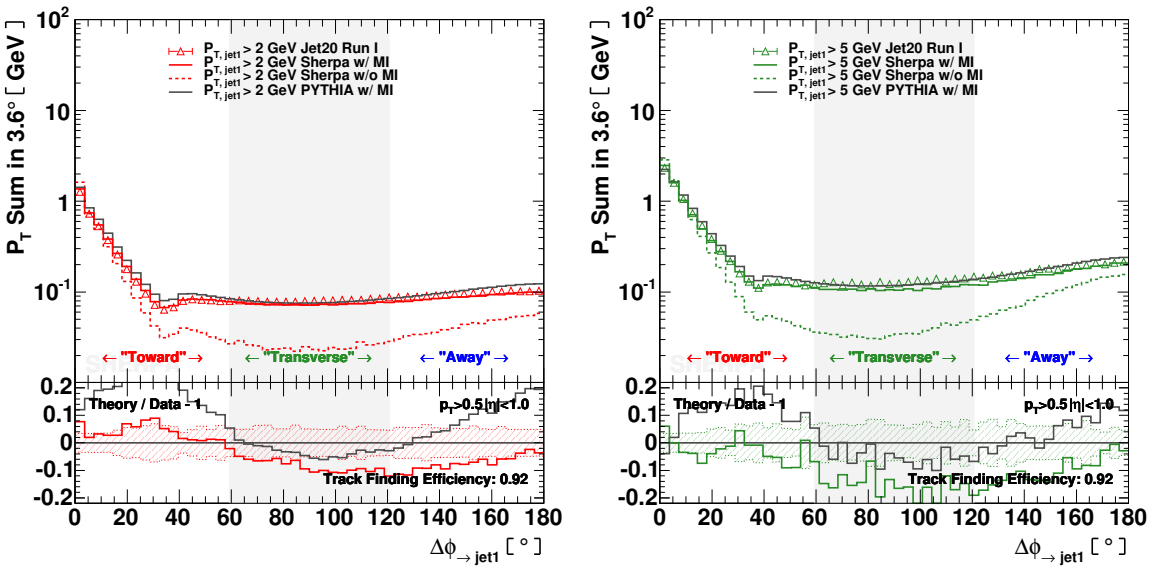


Fig. 4: Scalar P_T sum as a function of the azimuthal angle relative to the leading charged particle jet. The left figure shows results for $P_{T,jet1} > 2\,\mathrm{GeV}$, the right one displays results for $P_{T,jet1} > 5\,\mathrm{GeV}$.

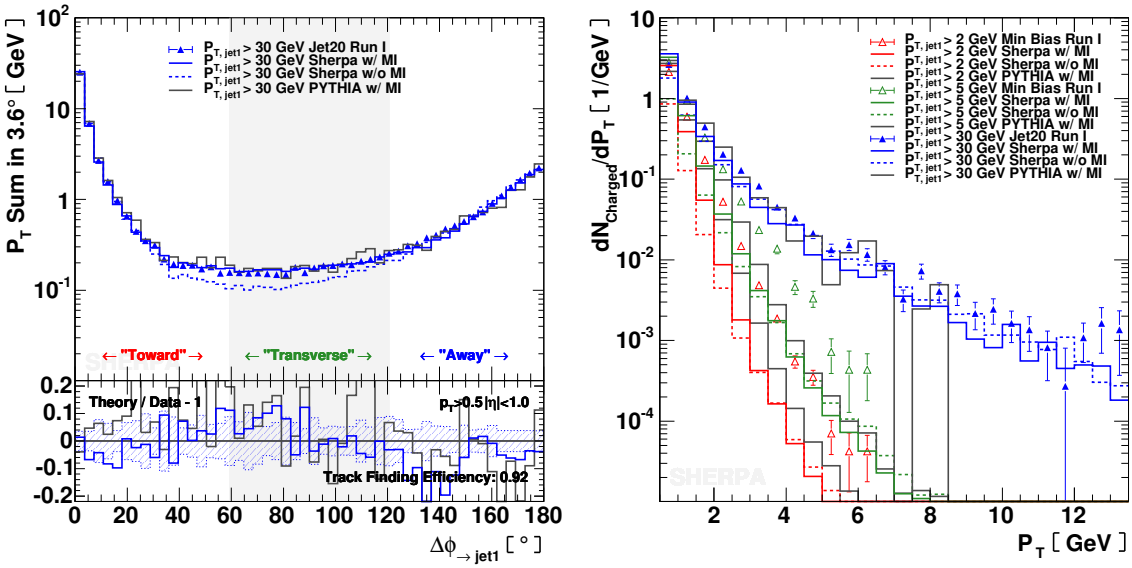


Fig. 5: Left: Scalar P_T sum as a function of the azimuthal angle relative to the leading charged particle jet for $P_{T,jet1} > 30\,\mathrm{GeV}$. Right: Charged particle multiplicity as a function of P_T in the "Transverse" region.

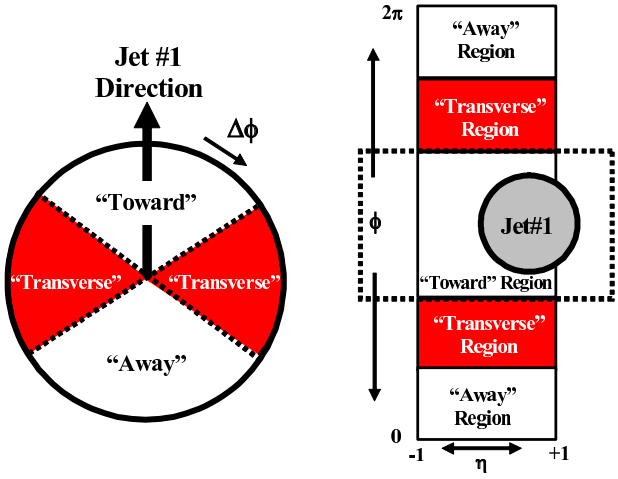


Fig. 6: Illustration of correlations in azimuthal angle ϕ relative to the direction of the leading jet (MidPoint, $R = 0.7$, $f_{merge} = 0.75$) in the event, jet#1. The angle $\Delta\phi = \phi - \phi_{\mathrm{jet1}}$ is the relative azimuthal angle between charged particles and the direction of jet#1. The "transverse" region is defined by $60° < |\Delta\phi| < 120°$ and $|\eta| < 1$. We examine charged particles in the range $p_T > 0.5\,\mathrm{GeV/c}$ and $|\eta| < 1$ and calorimeter towers with $|\eta| < 1$, but allow the leading jet to be in the region $|\eta(\mathrm{jet}\#1)| < 2$.

separate the "hard component" (initial and final-state radiation) from the "beam-beam remnant" component. MAX (MIN) refer to the "transverse" region containing largest (smallest) number of charged particles or to the region containing the largest (smallest) scalar $PT\mathrm{sum}$ of charged particles or the region containing the largest (smallest) scalar $ET\mathrm{sum}$ of particles. Since we will be studying regions in η-ϕ space with different areas, we will construct densities by dividing by the area. For example, the number density, $dN_{chg}/d\phi d\eta$, corresponds to the number of charged particles ($p_T > 0.5\,\mathrm{GeV/c}$) per unit η-ϕ the PTsum density, $dPT_{sum}/d\phi d\eta$, corresponds to the amount of charged particle ($p_T > 0.5\,\mathrm{GeV/c}$) scalar $PT\mathrm{sum}$ per unit η-ϕ, and the transverse energy density, $dET_{sum}/d\phi d\eta$, corresponds the amount of scalar $ET\mathrm{sum}$ of all particles per unit η-ϕ. One expects that the "transMAX" region will pick up the

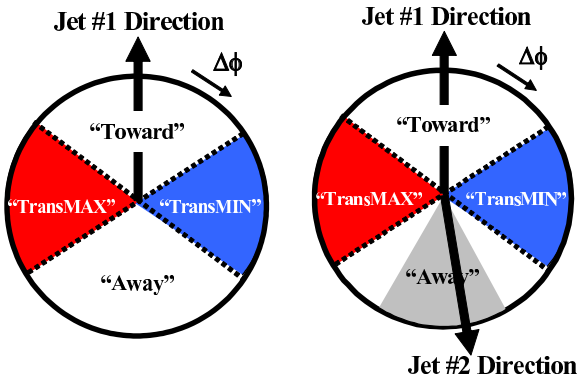


Fig. 7: Illustration of correlations in azimuthal angle ϕ relative to the direction of the leading jet (highest P_T jet) in the event, jet#1. The angle $\Delta\phi = \phi - \phi_{\rm jet\#1}$ is the relative azimuthal angle between charged particles and the direction of jet#1. On an event by event basis, we define "transMAX" ("transMIN") to be the maximum (minimum) of the two "transverse" regions, $60^\circ < \Delta\phi < 120^\circ$ and $60^\circ < -\Delta\phi < 120^\circ$. "transMAX" and "transMIN" each have an area in η-ϕ space of $\Delta\eta\Delta\phi = 4\pi/6$. The overall "transverse" region defined in Fig. 6 contains both the "transMAX" and the "transMIN" regions. Events in which there are no restrictions placed on the second and third highest p_T jets (jet#2 and jet#3) are referred to as "leading jet" events (*left*). Events with at least two jets with $p_T > 15$ GeV/c where the leading two jets are nearly "back-to-back" ($|\Delta\phi| > 150^\circ$) with $p_T(\text{jet\#2})/p_T(\text{jet\#1}) > 0.8$ and $p_T(\text{jet\#3}) < 15$ GeV/c are referred to as "back-to-back" events (*right*).

hardest initial or final-state radiation while both the "transMAX" and "transMIN" regions should receive "beam-beam remnant" contributions. Hence one expects the "transMIN" region to be more sensitive to the "beam-beam remnant" component of the "underlying event", while the "transMAX" minus the "transMIN" (i.e., "transDIF") is very sensitive to hard initial and final-state radiation. This idea, was first suggested by Bryan Webber and Pino Marchesini [25], and implemented in a paper by Jon Pumplin [26]. This was also studied by Valeria Tano in her CDF Run 1 analysis of maximum and minimum transverse cones [27].

Our previous Run 2 UE analysis [28] used JetClu to define jets and compared uncorrected data with PYTHIA Tune A [6] and HERWIG after detector simulation (i.e., CDFSIM). This analysis uses the MidPoint jet algorithm ($R = 0.7$, $f_{merge} = 0.75$) and corrects the observables to the particle level. The corrected observables are then compared with the QCD Monte-Carlo models at the particle level (i.e., generator level). The models includes PYTHIA Tune A, HERWIG and HERWIG with a tuned version of JIMMY [10]. In addition, for the first time we study the transverse energy density in the "transverse" region.

Fig. 8 compares the data on the density of charged particles and the charged $PT\text{sum}$ density in the "transverse" region corrected to the particle level for "leading jet" and "back-to-back" events with PYTHIA Tune A and HERWIG at the particle level. As expected, the "leading jet" and "back-to-back" events behave quite differently. For the "leading jet" case the "transMAX" densities rise with increasing $P_T(\text{jet\#1})$, while for the "back-to-back" case they fall with increasing $P_T(\text{jet\#1})$. The rise in the "leading jet" case is, of course, due to hard initial and final-state radiation, which has been suppressed in the "back-to-back" events. The "back-to-back" events allows a closer look at the "beam-beam remnant" and multiple parton scattering component of the UE. PYTHIA Tune A, which includes multiple parton interactions, does a better job of describing the data than HERWIG which does not have multiple parton interactions.

The "transMIN" densities are more sensitive to the "beam-beam remnant" and multiple parton interaction component of the "underlying event". The "back-to-back" data show a decrease in the "transMIN" densities with increasing $P_T(\text{jet\#1})$ which is described fairly well by PYTHIA Tune A (with multiple parton interactions) but not by HERWIG (without multiple parton interactions). The decrease

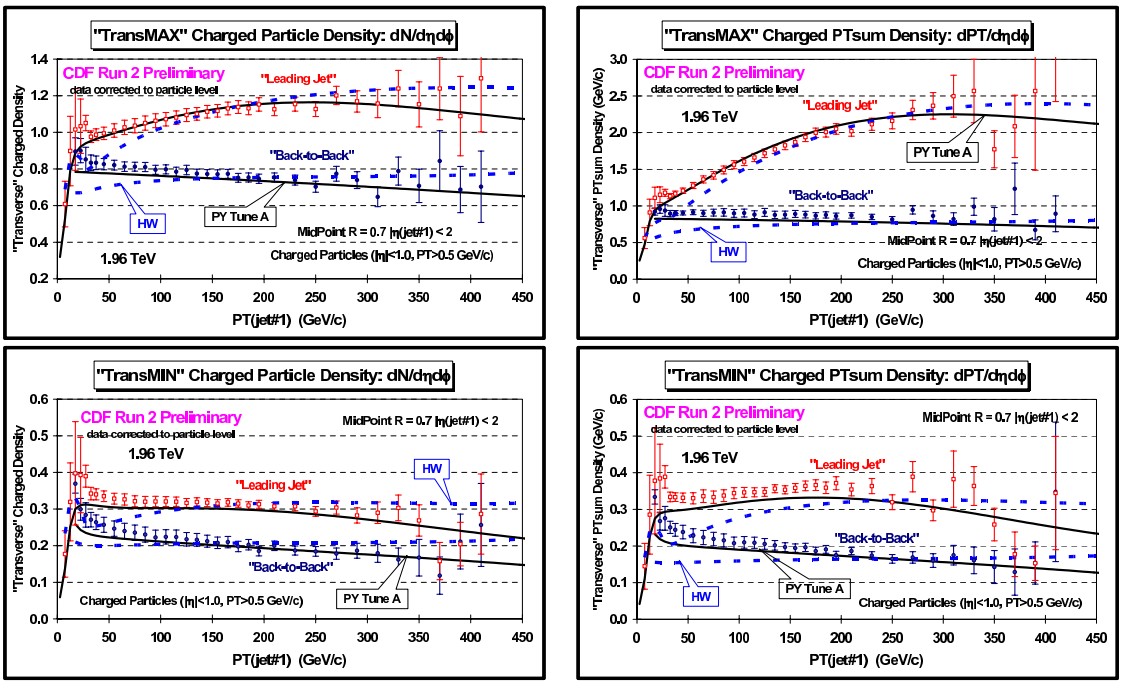


Fig. 8: Data at 1.96 TeV on (left) the density of charged particles $dN_{chg}/d\phi d\eta$ and (right) on the scalar $PT\mathrm{sum}$ density of charged particles, with $p_T > 0.5$ GeV/c and $|\eta| < 1$ in the "transMAX" region (*top*) and the "transMIN" region (*bottom*) for "leading jet" and "back-to-back" events defined in Fig. 7 as a function of the leading jet P_T compared with PYTHIA Tune A and HERWIG. The data are corrected to the particle level (with errors that include both the statistical error and the systematic uncertainty) and compared with the theory at the particle level (i.e., generator level).

of the "transMIN" densities with increasing $P_T(\mathrm{jet}\#1)$ for the "back-to-back" events is very interesting and might be due to a "saturation" of the multiple parton interactions at small impact parameter. Such an effect is included in PYTHIA Tune A but not in HERWIG (without multiple parton interactions).

Fig. 9(left) compares the data on average p_T of charged particles in the "transverse" region corrected to the particle level for "leading jet" and "back-to-back" events with PYTHIA Tune A and HERWIG at the particle level. Again the "leading jet" and "back-to-back" events behave quite differently.

Fig. 9(right) shows the data corrected to the particle level for the scalar $ET\mathrm{sum}$ density in the "transverse" region for "leading jet" and "back-to-back" events compared with PYTHIA Tune A and HERWIG. The scalar $ET\mathrm{sum}$ density has been corrected to correcpond to all particles (all p_T, $|\eta| < 1$). Neither PYTHIA Tune A nor HERWIG produce enough energy in the "transverse" region. HERWIG has more "soft" particles than PYTHIA Tune A and does slightly better in describing the energy density in the "transMAX" and "transMIN" regions.

Fig. 10(left) shows the difference of the "transMAX" and "transMIN" regions ("transDIF" = "transMAX" minus "transMIN") for "leading jet" and "back-to-back" events compared with PYTHIA Tune A and HERWIG. "TransDIF" is more sensitive to the hard scattering component of the UE (i.e., initial and final state radiation). Both PYTHIA Tune A and HERWIG underestimate the energy density in the "transMAX" and "transMIN" regions (see Fig. 9). However, they both fit the "transDIF" energy density. This indicates that the excess energy density seen in the data probably arises from the "soft" component of the UE (i.e., beam-beam remnants and/or multiple parton interactions).

JIMMY is a model of multiple parton interaction which can be combined with HERWIG to enhance the UE thereby improving the agreement with data. Fig. 10(right) and Fig. 11(left) show the energy density and charged $PT\mathrm{sum}$ density, respectively, in the "transMAX" and "transMIN" regions for "lead-

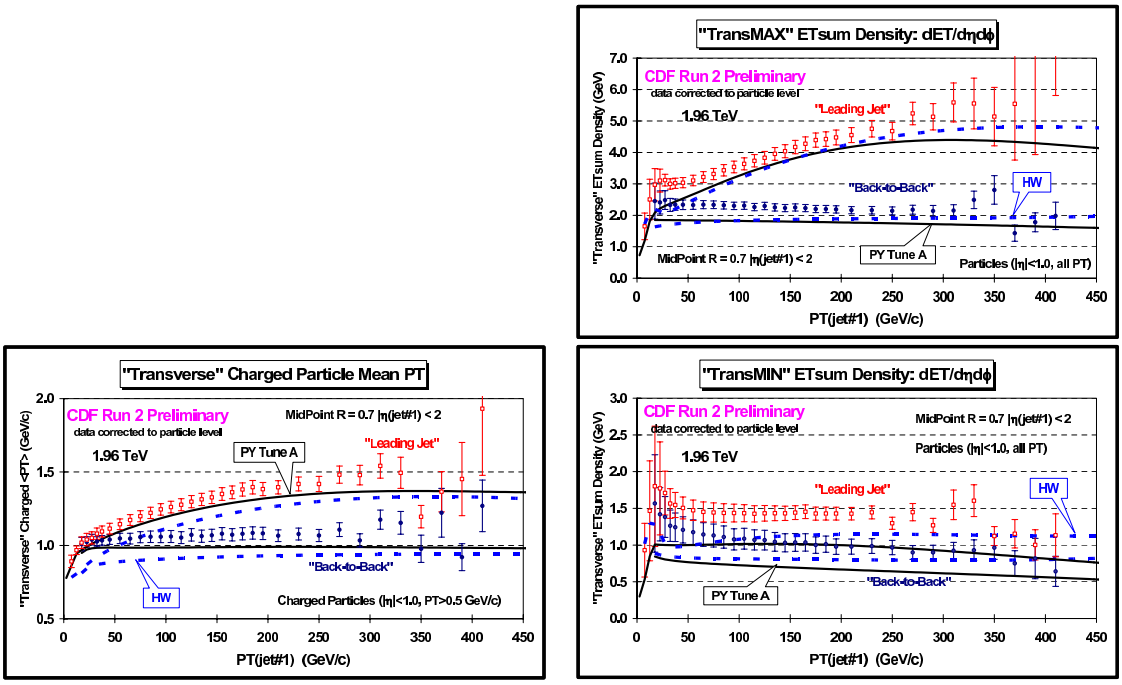


Fig. 9: On the left, data at 1.96 TeV on average transverse momentum, $\langle p_T \rangle$, of charged particles $|\eta| < 1$ in the with with $p_T > 0.5\,\mathrm{GeV/c}$ and $|\eta| < 1$ in the "transverse" region. On the right, scalar ETsum density, $dET_{sum}/d\phi d\eta$, for particles. with $p_T > 0.5\,\mathrm{GeV/c}$ and $|\eta| < 1$ in the "transMAX" region or the "transMIN" region. The "leading jet" and "back-to-back" events are defined in Fig. 7, and the data are shown as a function of the leading jet P_T and compared with PYTHIA Tune A and HERWIG. The data are corrected to the particle level (with errors that include both the statistical error and the systematic uncertainty) and compared with the theory at the particle level (i.e., generator level).

ing jet" and "back-to-back" events compared with PYTHIA Tune A and a tuned version of JIMMY˙JIMMY was tuned to fit the "transverse" energy density in "leading jet" events ($PTJIM = 3.25\,\mathrm{GeV/c}$). The default JIMMY ($PTJIM = 2.5\,\mathrm{GeV/c}$) produces too much energy and too much charged PTsum in the "transverse" region. Tuned JIMMY does a good job of fitting the energy and charged PTsum density in the "transverse" region (although it produces slightly too much charged PTsum at large $P_T(\mathrm{jet}\#1)$). However, the tuned JIMMY produces too many charged particles with $p_T > 0.5\,\mathrm{GeV/c}$ (see Fig. 11(right)). The particles produced by this tune of JIMMY are too soft. This can be seen clearly in Fig. 12 which shows the average charge particle p_T in the "transverse" region.

The goal of this analysis is to produce data on the UE that is corrected to the particle level so that it can be used to tune the QCD Monte-Carlo models without requiring CDF detector simulation. Comparing the corrected observables with PYTHIA Tune A and HERWIG at the particle level (i.e., generator level) leads to the same conclusions as we found when comparing the uncorrected data with the Monte-Carlo models after detector simulation [28]. PYTHIA Tune A (with multiple parton interactions) does a better job in describing the UE (i.e., "transverse" regions) for both "leading jet" and "back-to-back" events than does HERWIG (without multiple parton interactions). HERWIG does not have enough activity in the UE for $P_T(\mathrm{jet}\#1)$ less than about $150\,\mathrm{GeV/c}$, which was also observed in our published Run 1 analysis [19].

This analysis gives our first look at the energy in the UE (i.e., the "transverse" region). Neither PYTHIA Tune A nor HERWIG produce enough transverse energy in the "transverse" region. However, they both fit the "transDIF" energy density ("transMAX" minus "transMIN"). This indicates that the excess energy density seen in the data probably arises from the "soft" component of the UE (i.e., beam-

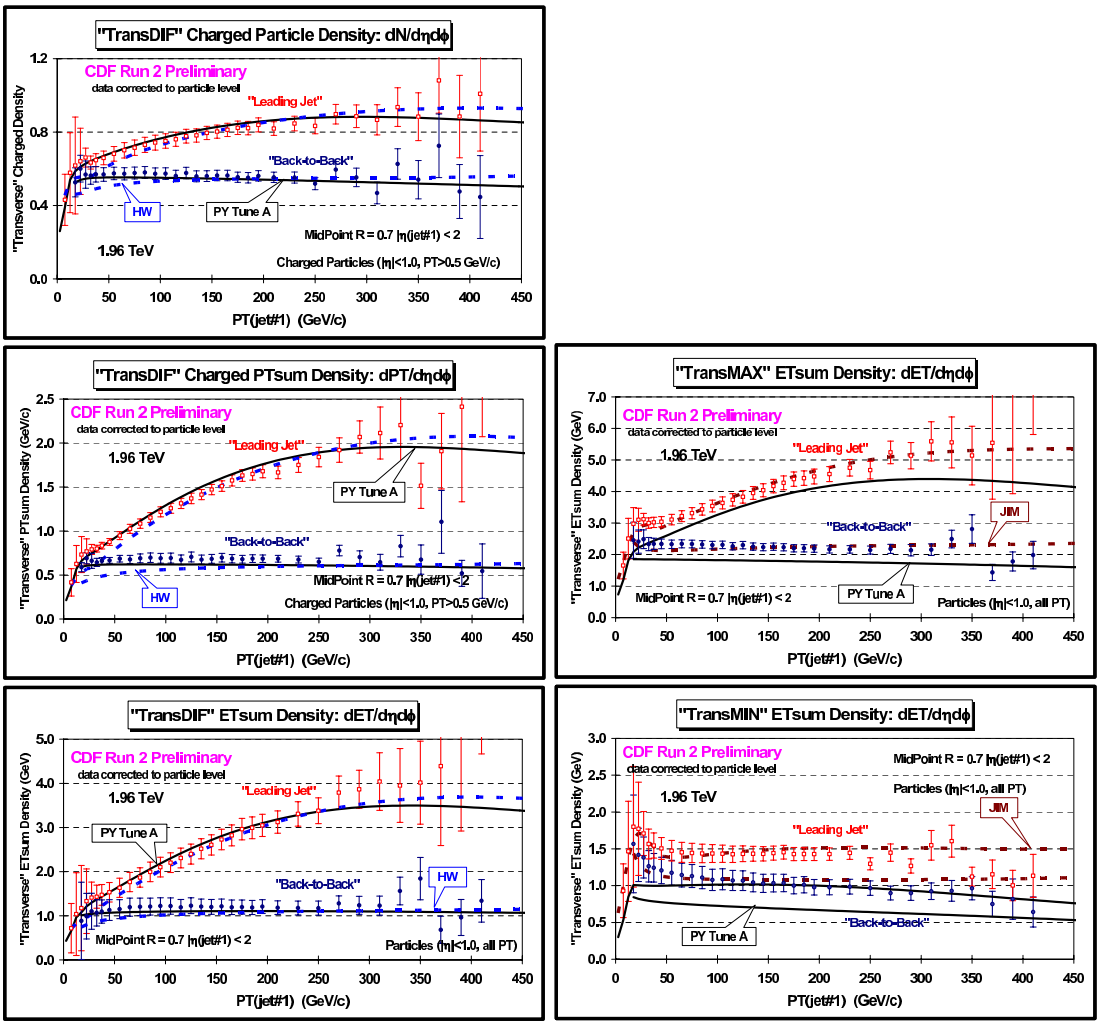


Fig. 10: Left: Data at 1.96 TeV on the difference of the "transMAX" and "transMIN" regions ("transDIF" = "transMAX"- "transMIN") for "leading jet" and "back-to-back" events defined in Fig. 7 as a function of the leading jet P_T compared with PYTHIA Tune A and HERWIG.
Right: Data on scalar ETsum density, $dET_{sum}/d\phi d\eta$, for particles with $|\eta| < 1$ in the "transMAX" region (*top*) and the "transMIN" region (*bottom*) for "leading jet" and "back-to-back" events defined in Fig. 7 as a function of the leading jet P_T compared with PYTHIA Tune A and tuned JIMMY. JIMMY was tuned to fit the "transverse" energy density in "leading jet" events ($PTJIM = 3.25$ GeV/c). The data are corrected to the particle level (with errors that include both the statistical error and the systematic uncertainty) and compared with the theory at the particle level (i.e., generator level).

beam remnants and/or multiple parton interactions). HERWIG has more "soft" particles than PYTHIA Tune A and does slightly better in describing the energy density in the "transMAX" and "transMIN" regions. Tuned JIMMY does a good job of fitting the energy and charged PTsum density in the "transverse" region (although it produces slightly too much charged PTsum at large $P_T(\text{jet}\#1)$). However, the tuned JIMMY produces too many charged particles with $p_T > 0.5$ GeV/c indicating that the particles produced by this tuned JIMMY are too soft.

In summary, we see an interesting dependence of the UE on the transverse momentum of the leading jet (i.e., the Q^2 of the hard scattering). For the "leading jet" case the "transMAX" densities rise with increasing $P_T(\text{jet}\#1)$, while for the "back-to-back" case they fall with increasing $P_T(\text{jet}\#1)$. The rise in the "leading jet" case is due to hard initial and final-state radiation with $p_T > 15$ GeV/c,

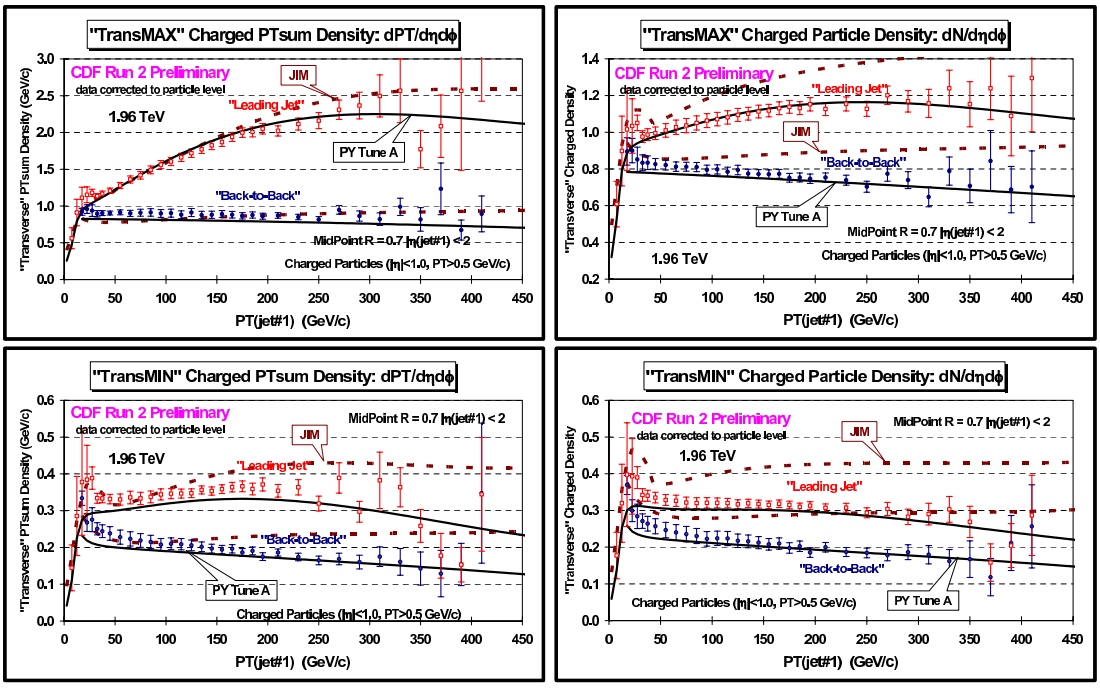


Fig. 11: Left: Data at 1.96 TeV on scalar $PT\mathrm{sum}$ density of charged particles, $dPT_{sum}/d\phi d\eta$, with $p_T > 0.5\,\mathrm{GeV/c}$ and $|\eta| < 1$ in the "transMAX" region (*top*) and the "transMIN" region (*bottom*) for "leading jet" and "back-to-back" events defined in Fig. 7 as a function of the leading jet P_T compared with PYTHIA Tune A and tuned JIMMY. JIMMY was tuned to fit the "transverse" energy density in "leading jet" events ($PTJIM = 3.25\,\mathrm{GeV/c}$). Right: Data on the density of charged particles, $dN_{chg}/d\phi d\eta$, with $p_T > 0.5\,\mathrm{GeV/c}$ and $|\eta| < 1$ in the "transMAX" region (*top*) and the "transMIN" region (*bottom*) for "leading jet" and "back-to-back" events defined in Fig. 2 as a function of the leading jet P_T compared with PYTHIA Tune A and tuned JIMMY. JIMMY was tuned to fit the "transverse" energy density in "leading jet" events ($PTJIM = 3.25\,\mathrm{GeV/c}$). The data are corrected to the particle level (with errors that include both the statistical error and the systematic uncertainty) and compared with the theory at the particle level (i.e., generator level).

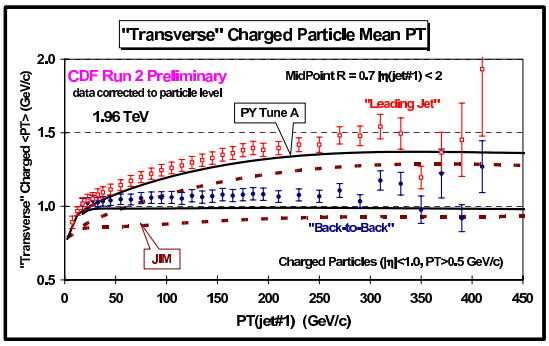


Fig. 12: Data at 1.96 TeV on average transverse momentum, $\langle p_T \rangle$, of charged particles with $p_T > 0.5\,\mathrm{GeV/c}$ and $|\eta| < 1$ in the "transverse" region for "leading jet" and "back-to-back" events defined in Fig. 7 as a function of the leading jet P_T compared with PYTHIA Tune A and tuned JIMMY. JIMMY was tuned to fit the "transverse" energy density in "leading jet" events ($PTJIM = 3.25\,\mathrm{GeV/c}$). The data are corrected to the particle level (with errors that include both the statistical error and the systematic uncertainty) and compared with the theory at the particle level (i.e., generator level).

which has been suppressed in the "back-to-back" events. The "back-to-back" data show a decrease in the "transMIN" densities with increasing $P_T(\mathrm{jet}\#1)$. The decrease of the "transMIN" densities with increasing $P_T(\mathrm{jet}\#1)$ for the "back-to-back" events is very interesting and might be due to a "saturation" of the multiple parton interactions at small impact parameter. Such an effect is included in PYTHIA Tune A (with multiple parton interactions) but not in HERWIG (without multiple parton interactions). PYTHIA Tune A does predict this decrease, while HERWIG shows an increase (due to increasing initial and final state radiation).

4 Extrapolation to LHC energies

The LHCb experiment [29] is designed to measure CP violation in the B-quark sector at the LHC and expand the current studies underway at the B-factories (BaBar, Belle) and at the Tevatron (CDF, D0). At $\sqrt{s}$=1.8 TeV, 28% of all of the primary produced B-mesons in $p\bar{p}$ collisions are produced in L=1 excited states [30]. These excited states decay via the emission of a charged hadron, allowing the possibility of same-side-tagging (SST) studies. As such, it is important to simulate the production of B mesons as accurately as possible.

The production of primary produced excited meson states are not included in the default PYTHIA [31] settings and including them increases the average multiplicity of an event. An attempt to reproduce the HFAG [32] values whilst retaining the spin counting rule for B** states has been made. This note covers a preliminary re-tuning [33] of PYTHIA v6.224 including these settings.

4.1 Method

The main parameter of the multiple-interaction model in PYTHIA v6.224 is the $\hat{p}_T^{\min}$ parameter, which defines the minimum transverse momentum of the parton-parton interactions. This effectively controls the number of parton-parton collisions and hence the average track multiplicity.

The charged particle density measured at $\eta = 0$ in the range of centre-of-mass energies, 52 GeV $< \sqrt{s} <$ 1800 GeV, [34] [35] is used to tune the $\hat{p}_T^{\min}$ parameter of PYTHIA. We define $\rho = \frac{1}{N_{ev}} \frac{dN_{ch}}{d\eta}|_{\eta=0}$ and measure ρ for a range of $\hat{p}_T^{\min}$ values at each $\sqrt{s}$. The quantity $\delta = \rho_{MC} - \rho_{Data}$ is plotted against $\hat{p}_T^{\min}$ and a linear fit performed. In Fig. 13, the re-tuned value of $\hat{p}_T^{\min}$ at $\sqrt{s} = 900$ GeV is taken to be the point at which the fit crosses the $\hat{p}_T^{\min}$ axis. To extrapolate $\hat{p}_T^{\min}$ to LHC energy, a fit is performed (Figure 14) using the form suggested by PYTHIA:

$$\hat{p}_T^{\min} = \hat{p}_T^{\min}(LHC)\Big(\frac{\sqrt{s}}{14TeV}\Big)^{2\epsilon} \tag{8}$$

4.2 Results

Extrapolating to $14\ TeV$ using the tuned values of $\hat{p}_T^{\min}(\sqrt{s})$ and (8), we obtain $\hat{p}_T^{\min}(LHC) = 3.34 \pm 0.13$, with $\epsilon = 0.079 \pm 0.0006$ with a corresponding central multiplicity of $\rho = 6.45 \pm 0.25$. Comparing the output of the re-tuned settings (dashed line) to the old LHCb settings (solid line), *Fig.* 15, 16 and 17, we find that the re-tuned settings produce a slightly lower multiplicity which affects the other distributions accordingly. Note: both the fragmentation parameters and the $\hat{p}_T^{\min}$ parameter affect the multiplicity of a generated event. This re-tuning method varies the $\hat{p}_T^{\min}$ parameter only i.e. it does not alter the fragmentation parameters in any fashion. Further investigations into re-tuning the fragmentation parameters using data from LEP are underway.

4.3 Conclusions

The central multiplicity values measured at CDF and UA5 are accurately reproduced using the re-tuned values for $\hat{p}_T^{\min}$ at several $\sqrt{s}$. An extrapolation of $\hat{p}_T^{\min}$ to LHC energies using a form implemented

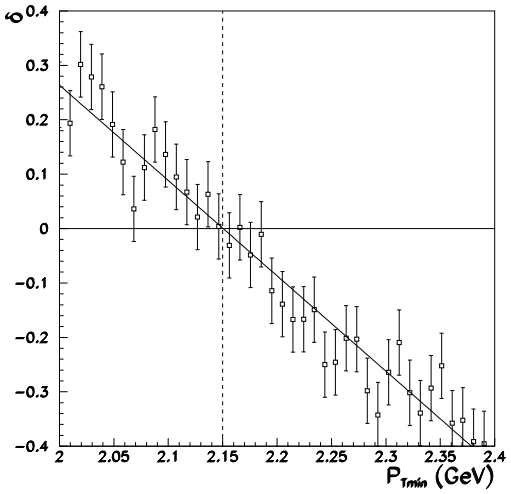


Fig. 13: Determining the value of $\hat{p}_T^{\min}(\sqrt{s} = 900GeV)$, the dashed line shows the point at which $|\delta|$ is minimised.

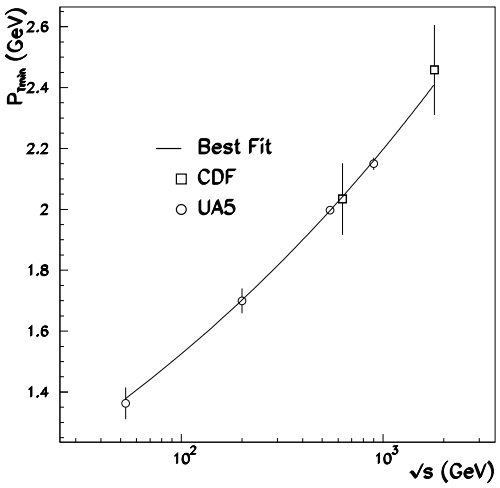


Fig. 14: The $\sqrt{s}$ dependance of $\hat{p}_T^{\min}$. The curve is the result of a fit assuming the functional form of (8).

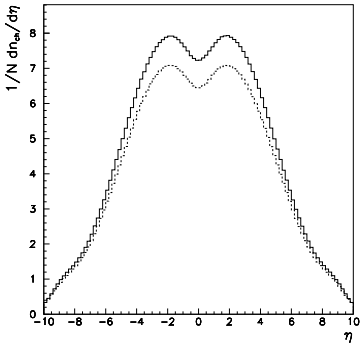


Fig. 15: η distribution at 14 TeV using the extrapolated value of $P_{T_{\rm Min}}$

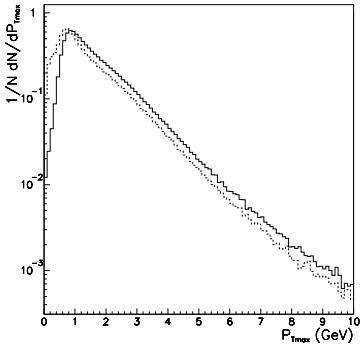


Fig. 16: $p_{\perp_{\max}}$ distribution in the LHCb acceptance

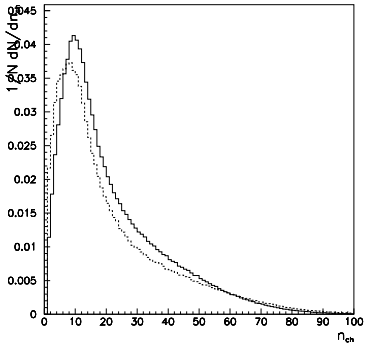


Fig. 17: Charged-stable multiplicity distribution in the LHCb acceptance.

in PYTHIA gives $\hat{p}_T^{\min}(LHC) = 3.34 \pm 0.13$, with $\epsilon = 0.079 \pm 0.0006$ with a corresponding central multiplicity of $\rho_{LHC} = 6.45 \pm 0.25$ in non-diffractive events.

5 Tuned models for the underlying event and minimum bias interactions

In this section we compare tuned MC generator models for the underlying event and minimum bias interactions. The aim of this study is to predict the event activity of minimum bias and the underlying event at the LHC. The models investigated correspond to tuned versions of PYTHIA, PHOJET and JIMMY.

5.1 Tuned models for the underlying event and minimum bias interactions

The starting point for the event generation in PYTHIA and JIMMY is the description of multiple hard interactions in the hadronic collision described in Section 2.1 (for PYTHIA 6.2), Section 2.2 for JIMMY and Section 2.4 for PHOJET.

Table 1: PYTHIA 6.214 default, ATLAS and CDF tune A parameters for minimum bias and the underlying event.

Default [31]	**ATLAS** [37]	**CDF tune A** [6]	**Comments**
MSTP(51)=7	MSTP(51)=7	MSTP(51)=7	CTEQ5L - selected p.d.f.
MSTP(81)=1	MSTP(81)=1	MSTP(81)=1	multiple interactions
MSTP(82)=1	MSTP(82)=4	MSTP(82)=4	complex scenario plus double Gaussian matter distribution
PARP(67)=1	PARP(67)=1	PARP(67)=4	parameter regulating initial state radiation
PARP(82)=1.9	PARP(82)=1.8	PARP(82)=2.0	$p_{t_{\min}}$ parameter
PARP(84)=0.2	PARP(84)=0.5	PARP(84)=0.4	hadronic core radius (only for MSTP(82)=4)
PARP(85)=0.33	PARP(85)=0.33	PARP(85)=0.9	probability for gluon production with colour connection to nearest neighbours
PARP(86)=0.66	PARP(86)=0.66	PARP(86)=0.95	probability to produce gluons either either as in PARP(85) or as a closed gluon loop
PARP(89)=1.0	PARP(89)=1.0	PARP(89)=1.8	energy scale (TeV) used to calculate $p_{t_{\min}}$
PARP(90)=0.16	PARP(90)=0.16	PARP(90)=0.25	power of the energy dependence of $p_{t_{\min}}$

PYTHIA and PHOJET have been shown to describe both minimum bias and underlying event data reasonably well when appropriately tuned [3, 6, 36, 37]. JIMMY is limited to the description of the underlying event; again, it has been shown capable of describing this rather well [38].

5.2 PYTHIA tunings

Several minimum bias and underlying event (UE) tunings for PYTHIA have been proposed in recent years. Ref. [37] describes how the current ATLAS tuning for PYTHIA was obtained after extensive comparisons to a variety of experimental measurements made at different colliding energies. Similar work has been done by the CDF Collaboration, although their PYTHIA tuning, CDF tune A [6], is primarily based on the description of the underlying event in jet events measured for $p\overline{p}$ at $\sqrt{s}$ = 1.8 TeV.

Table 1 displays the relevant parameters tuned to the data as proposed by the ATLAS [37] and CDF [6] collaborations. For the purpose of comparison, the corresponding default values [31] are also shown in the table.

5.3 PHOJET

The parameters used in PHOJET to describe minimum bias and the underlying event can be found in Ref. [22] and are currently set as default in PHOJET1.12, which is used in this study.

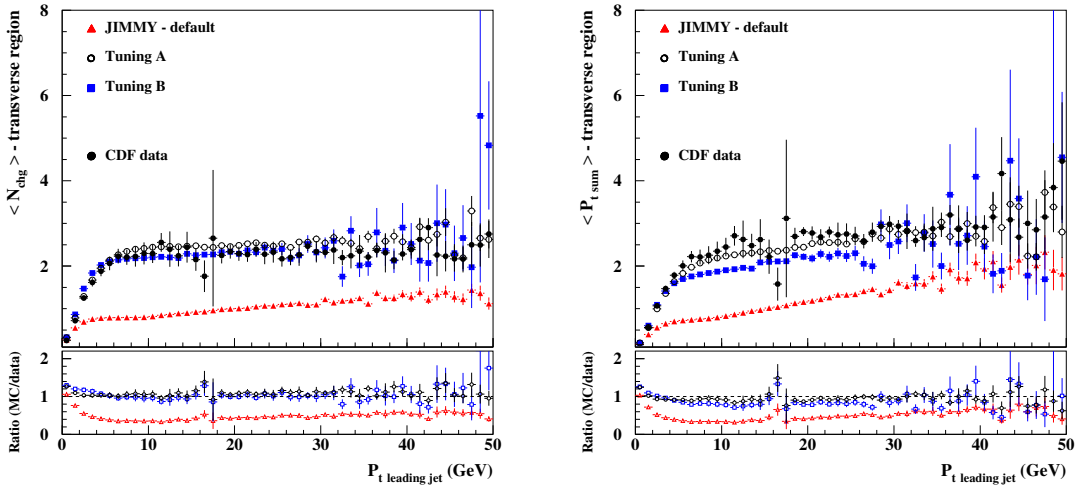
5.4 JIMMY tunings

We have tuned JIMMY to describe the UE as measured by CDF [19] and the resulting sets of parameters are shown in table 2. Figure 18 shows JIMMY predictions for the UE compared to CDF data for the average charged particle multiplicity (a) and the average p_t sum in the underlying event (b). In Fig.18 we compare JIMMY - default parameters to "Tuning A" and "Tuning B". Note that for the default parameters JIMMY does not give a correct description of the data. The other two distributions, generated with tuning A and B parameters, agree fairly well with the data.

In this study, JIMMY - tuning A and B will only be used to generate LHC predictions for the underlying event associated to jet events.

Table 2: JIMMY 4.1 default, tunings A and B parameters for the underlying event.

Default	Tuning A	Tuning B	Comments
JMUEO=1	JMUEO=0	JMUEO=0	multiparton interaction model
PTMIN=10.0	PTMIN=3.0	PTMIN=2.0	minimum p_T in hadronic jet production
PTJIM=3.0	–	–	minimum p_T of secondary scatters when JMUEO=1 or 2
JMRAD(73)=0.71	JMRAD(73)=2.13	JMRAD(73)=0.71	inverse proton radius squared
PRSOF=1.0	PRSOF=0.0	PRSOF=0.0	probability of a soft underlying event



(a) (b)

Fig. 18: JIMMY predictions for the UE compared to CDF data. (a) Average charged particles multiplicity in the UE and (b) average p_t sum in the UE.

5.5 Minimum bias interactions at the LHC

Throughout this report, minimum bias events will be associated with non-single diffractive inelastic interactions, following the experimental trend (see Ref. [37] and references therein).

For LHC collisions (pp collisions at $\sqrt{s} = 14$ TeV) the minimum bias cross-section estimated by PYTHIA 6.214, regardless of which tuning is used, is $\sigma_{nsd} = 65.7$ mb while PHOJET1.12 predicts $\sigma_{nsd} = 73.8$ mb, 12.3% greater than the former. Hence, for the same luminosity PHOJET1.12 generates more minimum bias pp collisions than PYTHIA 6.214 - tuned. We shall however, focus on the general properties per pp collision not weighted by cross-sections. The results per pp collision can later be easily scaled by the cross-section and luminosity.

Figure 19(a) shows charged particle density distributions in pseudorapidity for minimum bias pp collisions at $\sqrt{s}$ = 14 TeV generated by PHOJET1.12 and PYTHIA 6.214 - ATLAS and CDF tune A. The charged particle density generated by PHOJET1.12 and PYTHIA 6.214 - CDF tune A and ATLAS at $\eta = 0$ is 5.1, 5.3 and 6.8, respectively. Contrasting to the agreement seen in previous studies for $p\overline{p}$ collisions at $\sqrt{s}$ = 200 GeV, 546 GeV, 900 GeV and 1.8 TeV in Ref. [37], at the LHC PYTHIA 6.214 - ATLAS generates $\sim 25\%$ more charged particle density in the central region than PYTHIA 6.214 - CDF tune A and PHOJET1.12.

Compared to the charged particle density $dN_{ch}/d\eta$ measured by the CDF experiment at 1.8 TeV [39], PYTHIA 6.214 - ATLAS indicates a plateau rise of $\sim 70\%$ at the LHC in the central region while

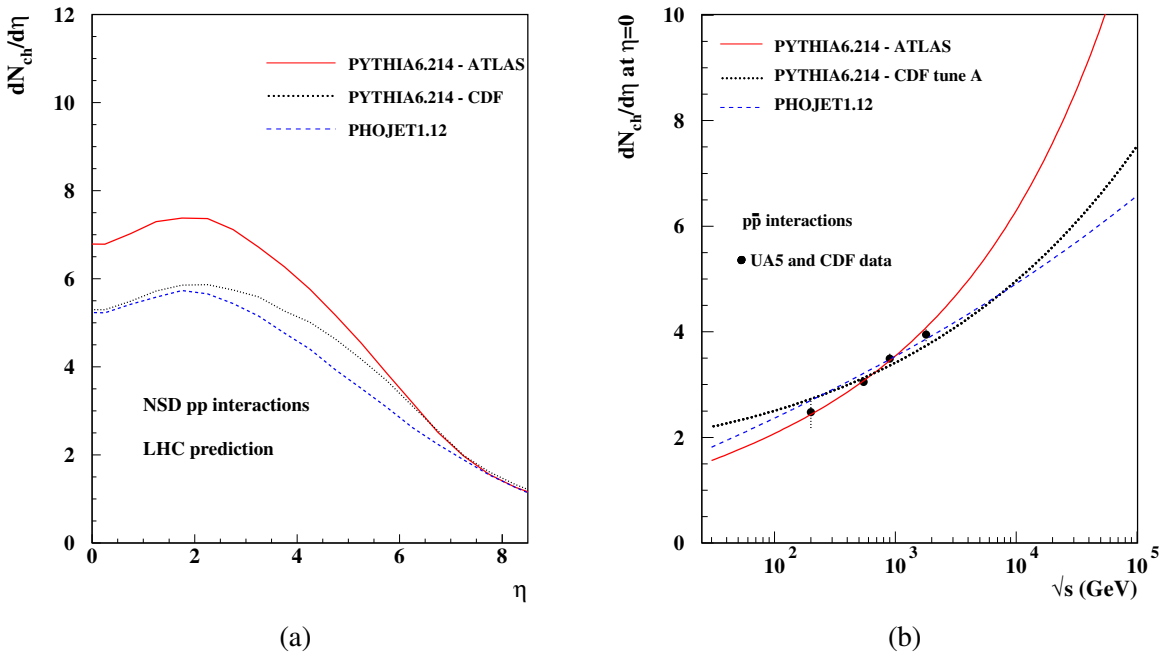


Fig. 19: (a) Charged particle density distributions, $dN_{ch}/d\eta$, for NSD pp collisions at $\sqrt{s}$ = 14 TeV. (b) $dN_{ch}/d\eta$ at $\eta = 0$ for a wide range of $\sqrt{s}$. Predictions generated by PYTHIA 6.214, ATLAS and CDF tune A and PHOJET1.12.

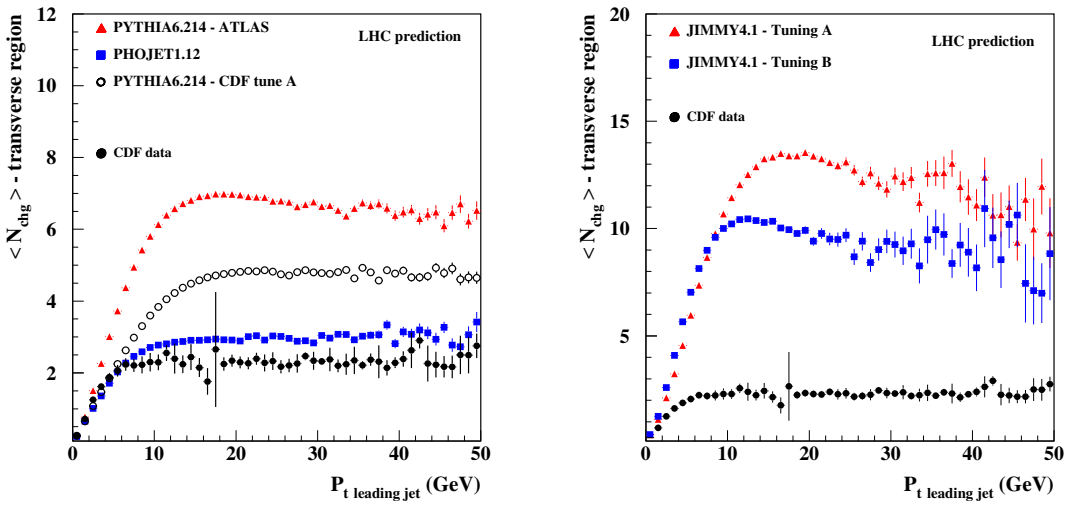
PHOJET1.12 and PYTHIA 6.214 - CDF tune A suggest a smaller rise of $\sim 35\%$.

Figure 19(b) displays $dN_{ch}/d\eta$ at $\eta = 0$ plotted as a function of $\sqrt{s}$. For centre-of-mass energies greater than $\sim$ 1 TeV, the multiparton interaction model employed by PYTHIA and the DPM used by PHOJET lead to multiplicity distributions with different rates of increase with the energy. PYTHIA suggests a rise dominated by the $\ln^2(s)$ term while PHOJET predicts that the dominant term gives a ln(s) rise for $dN_{ch}/d\eta$ at $\eta = 0$. The ATLAS tuning for PYTHIA gives a steeper rise than CDF tune A and PHOJET (Fig. 19(b)) indicating a faster increase in the event activity at the partonic level in the ATLAS tuning when compared to CDF tune A and PHOJET. The average charged particle multiplicity in LHC minimum bias collisions, $< n_{ch} >$, is 69.6, 77.5 and 91.0 charged particles as predicted by PHOJET1.12, PYTHIA 6.214 - CDF tune A and ATLAS, respectively.

The $< p_t >$ at $\eta = 0$ for charged particles in LHC minimum bias collisions predicted by PHOJET1.12 and PYTHIA 6.214 - ATLAS and CDF tune A models is 0.64 GeV, 0.67 GeV and 0.55 GeV, respectively. Generating less particles in an average minimum bias collision at the LHC, PHOJET1.12 predicts that the average p_t per particle at $\eta = 0$ is greater (or harder) than the corresponding prediction from PYTHIA 6.214 - ATLAS. However, amongst the three models, PYTHIA 6.214 - CDF tune A gives the hardest $< p_t >$ at $\eta = 0$. The main reason for this is the increased contribution of harder parton showers used to make the model agree with the p_t spectrum of particles in the UE, and obtained by setting PARP(67)=4 [6].

5.6 The underlying event

Based on CDF measurements, we shall use their definition for the UE, i.e., the angular region in ϕ which is transverse to the leading charged particle jet as described in Section 3 and shown in Fig. 6. Figure 20(a) displays PYTHIA 6.214 — ATLAS and CDF tune A, and PHOJET1.12 predictions for the average particle multiplicity in the UE for pp collisions at the LHC (charged particles with $p_T > 0.5$ GeV and $|\eta| < 1$). The distributions generated by the three models are fundamentally different. Except for events



(a) (b)

Fig. 20: (a) PYTHIA 6.214 (ATLAS and CDF tune A), PHOJET1.12 and (b) JIMMY 4.1 (tunings A and B) predictions for the average multiplicity in the UE for LHC pp collisions.

with $p_{t_{\rm ljet}} \lesssim 3$ GeV, PYTHIA 6.214 — ATLAS generates greater multiplicity in the UE than the other models shown in Fig. 20(a).

A close inspection of predictions for the UE given in Fig. 20(a), shows that the average multiplicity in the UE for $P_{t_{\rm ljet}} > 10$ GeV reaches a plateau at ~ 6.5 charged particles according to PYTHIA 6.214 - ATLAS, ~ 5 for CDF tune A and ~ 3.0 according to PHOJET1.12. Compared to the underlying event distributions measured by CDF at 1.8 TeV, PYTHIA 6.214 - ATLAS indicates a plateau rise of $\sim 200\%$ at the LHC while PYTHIA 6.214 - CDF tune A predicts a rise of $\sim 100\%$ and PHOJET1.12 suggests a much smaller rise of $\sim 40\%$.

In Fig. 20(b) we show JIMMY 4.1 - Tuning A and B predictions for the average particle multiplicity in the UE for LHC collisions. The average multiplicity in the UE for $P_{t_{\rm ljet}} > 10$ GeV reaches a plateau at ~ 12 charged particles according to JIMMY 4.1 - Tuning A, and ~ 9.0 according to JIMMY 4.1 - Tuning B. Note that, for both JIMMY tunings, the plateau rise for the average multiplicity in the UE is much greater than the ones predicted by any of the PYTHIA tunings or by PHOJET as shown in Figs. 20(a) and (b). Once again, compared to the underlying event distributions measured by CDF at 1.8 TeV, JIMMY 4.1 - Tuning A indicates a five-fold plateau rise at the LHC while JIMMY 4.1 - Tuning B - CDF suggests a four-fold rise.

5.7 Conclusion

The minimum bias and underlying event predictions for the LHC generated by models which have been tuned to the available data have been compared. In previous studies, these models have been shown to be able to describe the data distributions for these two classes of interactions. However, in this article, it has been shown that for the models detailed in tables 1 and 2, there can be dramatic disagreements in their predictions at LHC energies. This is especially evident in the distributions for the average multiplicity in the UE (Fig. 20) where, for example, PHOJET1.12 predicts that the distribution's plateau will be at ~ 3 charged particles while JIMMY 4.1 - Tuning A predicts for the same distribution, a plateau at ~ 12.

Even though models tuned to the data have been used in this study, uncertainties in LHC predictions for minimum bias and the underlying event are still considerable. Improved models for the soft component of hadronic collisions are needed as well as more experimental information which may be

used to tune current models. Future studies should focus on tuning the energy dependence for the event activity in both minimum bias and the underlying event, which at the moment seems to be one of the least understood aspects of all the models investigated in this study.

6 Can the final state at LHC be determined from ep data at HERA?

6.1 Jets and $E_\perp$-flow

A phenomenological fit for a soft-cutoff, $\hat{p}_T^{\min}$, and an extrapolation to LHC energies, was discussed in sections 4.1 and 5.2. However, in the $k_\perp$-factorization formalism the soft divergence is avoided, and it is possible to predict minijets and $E_\perp$-flow from HERA data alone. Thus it is not necessary to rely on a purely phenomenological fit using $\mathrm{p\bar{p}}$ collision data. This gives a better dynamical insight, and avoids the uncertainties associated with the extrapolation to higher energies.

High $p_\perp$ jets are well described by conventional *collinear factorization*, but in this formalism the minijet cross section diverges, $\sigma_{jet} \propto 1/p_\perp^4$. This implies that the total $E_\perp$ also diverges, and therefore a cutoff $\hat{p}_T^{\min}$ is needed. Fits to data give $\hat{p}_T^{\min} \sim 2$ GeV growing with energy [8,9]. There is no theoretical basis for the extrapolation of $\hat{p}_T^{\min}$ from the Tevatron to LHC, which induces an element of uncertainty in the predictions for LHC.

In the *$k_\perp$-factorization* formalism the off shell matrix element for the hard subcollision $k_1 + k_2 \to q_1 + q_2$ does not blow up, when the momentum exchange $k_\perp^2$ is smaller than the incoming virtualities $k_{\perp 1}^2$ and $k_{\perp 2}^2$. The unintegrated structure functions $\mathcal{F}(x, k_\perp^2, Q^2)$ are also suppressed for small $k_\perp$, and as a result the total $E_\perp$ is not divergent but stays finite. An "effective cutoff" increases with energy, but the increase is less steep for larger energies [40].

At high energy σ_{jet} is larger than σ_{tot}, which implies that there usually are *multiple hard subcollisions* in a single event. The experimental evidence for multiple collisions has been discussed in previous sections. It includes multijet events, forward-backward correlations, the pedestal effect, and associated particles in jet events. The data also indicate that the hard subcollisions are not independent. Central collisions contain more, and peripheral collisions fewer, minijets, and the results are well described by a double Gaussian distribution in impact parameter, as suggested in ref. [3].

At high energies the pdfs needed to calculate the minijet cross section have to be evaluated in the BFKL domain of small x and low $k_\perp$. This implies that non-$k_\perp$-ordered parton chains are important. For a γ^*p collision a single local $k_\perp$-maximum corresponds to a resolved photon interaction. Similarly several local maxima in a single chain correspond to correlated hard subcollisions.

In the BFKL formalism the gluon links in the t-channel correspond to reggeized gluons, which means that soft emissions are compensated by virtual corrections. These soft emissions *do not* contribute to the parton distributions or total cross sections, but they *do* contribute to the properties of final states, and should then be added with Sudakov form factors. The CCFM model [41, 42] interpolates between DGLAP and BFKL. Here some soft emissions are included in the initial state radiation, which implies that they must be suppressed by non-eikonal form factors. The Linked Dipole Chain (LDC) model [43] is a reformulation and generalization of CCFM, in which more emissions are treated as final state emissions, in closer agreement with the BFKL picture. In the LDC formalism the chain formed by the initial state radiation is *fully symmetric* with respect to the photon end and the proton end of the ladder. This symmetry implies that the formalism is also directly applicable to hadron-hadron collisions. Thus a fit to DIS data will also give the cross section for a *parton chain* in pp collisions [44].

A potential problem is due to the fact that with a running α_s, the enhancement of small $k_\perp$ implies that the result depends on a necessary cutoff Q_0. Good fits to DIS data are possible with different Q_0, if the input distribution $f_0(x, Q_0^2)$ is adjusted accordingly. However, although a larger cutoff gives fewer hard chains, it also implies a larger number of soft chains, in which no link has a $k_\perp$ larger than Q_0. Thus the total number of chains in pp scattering is independent of Q_0, and therefore well determined by the fit to DIS data.

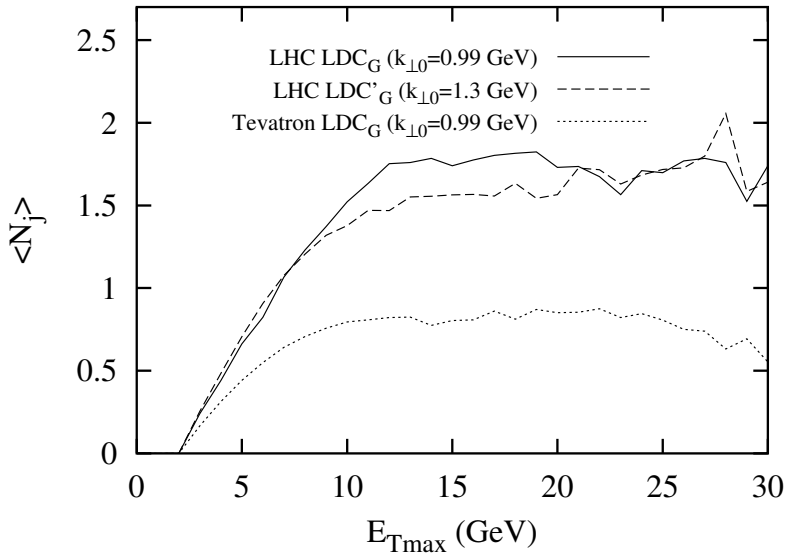


Fig. 21: The average number of minijets per event in the "minimum azimuth region", as a function of transverse energy of the trigger jet, $E_{\perp max}$. The figure shows the result for 1.8 TeV and for LHC. The two LHC curves correspond to different values for Q_0, showing the stability with respect to the soft cutoff.

When the fit to HERA data in this way is applied to p$\bar{\text{p}}$ scattering at the Tevatron, the predictions for *e.g.* jet multiplicity and the pedestal effect are very close to CDF's tune A, described in Section 3. The result is insensitive to the soft cutoff Q_0, which implies that the extrapolation to LHC energies is stable, and does not depend on an uncertain extrapolation of the low-$p_\perp$ cutoff needed in a collinear formalism. As an example fig. 21 shows a prediction for the average number of minijets per event within 60° in azimuth perpendicular to a trigger jet, on the side with minimum activity.

As the LDC model is fully symmetric with respect to an interchange of the projectile and the target, the parton chains have to combine at one end at the same rate as they multiply at the other. Therefore the formalism should be suitable for studies of gluon recombination and *saturation*. This work is in progress, and some preliminary results from combining the LDC model with Mueller's dipole formulation in transverse coordinate space [45–47] are presented in ref. [48].

6.2 Hadron multiplicities

The hadron multiplicity is much more sensitive to non-perturbative effects. This implies larger uncertainties, and models differ by factors 3-4 in their predictions for LHC (see Section 5). The CDF data also show that the data are best fitted if colours rearrange so that secondary hard scatterings give minimum extra string length, i.e. minimum extra multiplicity. This is very different from the case in e^+e^- annihilation.

In pp collisions the multiplicity of final state hadrons depends very sensitively on the colour connections between the produced partons. This implies that the result depends on soft non-perturbative effects. Multiple interactions are related to multiple pomeron exchange, which is expected to obey the Abramovskyĭ-Gribov-Kancheli cutting rules [49]. These rules are derived for a multiperipheral model, but a multiperipheral chain has important similarities with a gluonic chain. An essential feature is the dominance of small momentum exchanges at each vertex. The colour structure of QCD gives, however, some extra complications as discussed by J. Bartels (see the contribution by Bartels to working group 4).

The pomeron is identified by two gluon exchange, and multiple chains correspond to multi-pomeron exchange. For the example of two pomeron exchange, the AGK rules give the relative weights

$1 : -4 : 2$ for cutting 0, 1 or 2 pomerons. These ratios imply that the two-pomeron diagram contributes to the multiplicity *fluctuations*, but has no effect on the number of produced particles, determined by $\sum n\sigma_n$. This result can also be generalized to the exchange of more pomerons.

Similar cutting rules apply to a diagram with two pomerons attached to one proton and one pomeron to the other, connected by a central triple-pomeron coupling. In ref. [49] this and similar diagrams are, however, expected to give smaller contributions.

A hard gg $\rightarrow$ gg subcollision will imply that the two proton remnants carry colour octet charges. This is expected to give two colour triplet strings, or two cluster chains, connecting the two remnants and the two final state gluons. In the string model the strings are stretched between the remnants, with the gluons acting as kinks on the strings. These kinks can either be on different strings or both on the first or both on the second string, with equal probabilities for the three possibilities (see ref. [50]). Including initial state radiation will give extra kinks, which due to colour coherence will be connected so as to result in minimal extra string length.

Multiple collisions with two independent gg $\rightarrow$ gg scatterings would be expected to correspond to two cut pomerons, with four triplet strings stretched between the proton remnants. This would give approximately a doubled multiplicity, in accordance with the AGK cutting rules. However, the CDF data show that this is *far from reality*.

CDF's successful tune A [6] is a fit using an early PYTHIA version. Already in the analysis in ref. [3] it was realized that four strings would give too high multiplicity. Therefore in this early PYTHIA version there are three possible string connections for a secondary hard subcollision. 1) An extra closed string loop between the two final state gluons. 2) A single string between the scattered partons, which are then treated as a $q\bar{q}$ system. 3) The new hard gluons are inserted as extra kinks among the initial state radiations, in a way which corresponds to minimum extra string length. In the successful tune A the last possibility is chosen in 90% of the cases, which corresponds to *minimal extra multiplicity*. The default PYTHIA tune, which contained equal probabilities for the three cases, does not give a good fit. A more advanced treatment of pp collisions [8, 9] is implemented in a new PYTHIA version (6.3) [2] (see Section 2.1). This model does, however, not work as well as Field's tune A of the older model.

Consequently two independent hard collisions do not correspond to two cut pomeron ladders stretched between the proton remnants. It also does not correspond to a cut pomeron loop in the centre. Instead it looks like a single ladder, with a higher density of gluon rungs in the central region.

How can this be understood? It raises a set of important questions: What does it imply for the AGK rules and the diffractive gap survival probability? Do rescattering and unitarity constraints (and AGK) work in the initial perturbative phase? If so, does this correspond to an initial hard collision inside a confining bag, with the final state partons colour connected in a later non-perturbative phase?

We can compare with the situation in e^+e^--annihilation. If two gluons are emitted from the quark or antiquark legs, these gluons form a colour singlet with probability $\sim 1/N_c^2$. They could then hadronize as a separate system. Analyses of data from LEP indicate that such isolated systems are suppressed even more than by a factor $\sim 1/N_c^2$.

In conclusion we have following important questions:

- Why do the strings make the shortest connections in $\approx 100\%$ in pp and almost never in e^+e^-?
- How do multiplicity fluctuations and the relation diffraction *diffraction* and *high multiplicity events* reflect features of AGK in ep, γp, and pp?
- Do unitarity effects and AGK cutting rules work as expected in an initial perturbative phase, and the colours recombine in a subsequent nonperturbative soft phase?
- Or is the pomeron a much more complicated phenomenon than the simple ladder envisaged by Abramovskyĭ-Gribov-Kancheli?

7 Conclusions and the potential for HERA data

This was a very active area of discussion during the workshop. In fact, the area remains so active that firm conclusions are hard to make, and likely to be superceded on a very short timescale. Nevertheless there are some things which do seem clear.

- The underlying event is clearly an topic of substantial importance for the LHC.
- The dominant input data for understanding the underlying event comes at present from the Tevatron, with HERA data primarily featuring indirectly, though importantly, via the parton densities.
- The data strongly indicate that multiple hard scatters are required to adequately describe the final state in high energy hadron collisions.
- The UE depends on the measurement being made as demonstrated by difference between the UE in the CDF leading jet and back-to-back jet analysis.
- The colour structure of the final parton state is an unsolved problem. The CDF data indicate that 'short strings' are strongly favoured.
- There are large uncertainties associated with extrapolating the available models to LHC energies.

As far as the future impact of HERA data on this area goes, some ideas have been discussed in the previous section. In addition, it is worth noting that most of the models discussed here have also been used in high energy photoproduction at HERA [51], where they also improve the description of the data. No study comparable to those carried out at pp or $\mathrm{p\bar{p}}$ experiments is currently available. The benefits of such a study would be that (a) HERA could add another series of points in energy (around 200 GeV) to help pin down the energy dependence of the underlying event, (b) it is possible to select regions of phase space where resolved (i.e., hadronic) or direct (i.e., pointlike) photons dominate, thus effectively switching on or off the photon PDF (and thus presumably multiparton interactions) and allowing comparison between the two cases, (c) the photon is a new particle with which the physics assumptions of underlying event models can be confronted. The last of these points however also implies that a slew of new parameters will be introduced, and one may learn more about the photon this way than about underlying events themselves. Either way, it is to be hoped that such a study will be carried out before HERA finishes and LHC switches on.

References

[1] T. Sjöstrand *et al.*, Comput. Phys. Commun. **135**, 238 (2001). hep-ph/0010017.

[2] T. Sjöstrand *et al.*, LU TP 03-38 (2003). hep-ph/0308153.

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

[4] ZEUS Collaboration, S. Chekanov *et al.*, Eur. Phys. J. **C21**, 443 (2001);
H1 Collaboration, C. Adloff *et al.*, Eur. Phys. J. **C21**, 33 (2001).

[5] J. Dischler and T. Sjöstrand, Eur. Phys. J. direct **C3**, 2 (2001). hep-ph/0011282.

[6] R. Field, *Min-bias and the underlying event at the tevatron and the lhc*.
http://www.phys.ufl.edu/∼rfield/cdf/tunes/py_tuneA.html.

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

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

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

[10] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, Z. Phys. C **72**, 637 (1996).
(Hep-ph/9601371).

[11] J. M. Butterworth and M. H.Seymour, In preparation see
http://www.cedar.ac.uk/hepforge/jimmy.

[12] S. Höche and F. Krauss, in preparation.

[13] R. Kuhn, F. Krauss, B. Ivanyi, and G. Soff, Comput. Phys. Commun. **134**, 223 (2001). hep-ph/0004270.
[14] F. Krauss, A. Schälicke, and G. Soff, hep-ph/0503087.
[15] S. Catani *et al.*, JHEP **11**, 063 (2001). hep-ph/0109231.
[16] F. Krauss, JHEP **08**, 015 (2002). hep-ph/0205283.
[17] F. Krauss *et al.*, Phys. Rev. **D70**, 114009 (2004). hep-ph/0409106.
[18] A. Schälicke and F. Krauss, JHEP **07**, 018 (2005). hep-ph/0503281.
[19] CDF Collaboration, T. Affolder *et al.*, Phys. Rev. **D65**, 092002 (2002).
[20] R. D. Field, http://www.phys.ufl.edu/∼rfield/cdf/chgjet/chgjet_intro.html.
[21] A. Capella *et al.*, Phys. Rep. **236**, 225 (1994).
[22] R. Engel, *PHOJET manual (program version 1.05)*, June 1996.
[23] R. Field, Int. J. Mod. Phys. **A16S1A**, 250 (2001).
[24] J. Huston, Int. J. Mod. Phys. **A16S1A**, 219 (2001).
[25] G. Marchesini and B. R. Weber, Phys. Rev **D38**, 3419 (1988).
[26] J. Pumplin, Phys. Rev. **D57**, 5787 (1998).
[27] D. Acosta *et al.*, Phys. Rev. **D70**, 072002 (2004).
[28] R. Field, Acta Phys. Polon. **B36**, 167 (2005).
[29] LHCb Collaboration. CERN-LHCC-2003-030.
[30] CDF Collaboration, T. Affolder *et al.*, Phys. Rev. **D64**, 072002 (2001).
[31] T. Sjöstrand, L. Lönnblad, and S. Mrenna, *Pythia 6.2: Physics and manual*. Preprint hep-ph/0108264, 2001.
[32] Heavy Flavor Averaging Group (HFAG) (2005). hep-ex/0505100.
[33] LHCb Collaboration, P. Bartalini *et al.* (1999).
[34] UA5 Collaboration, G. J. Alner *et al.*, Z. Phys. **C33**, 1 (1986).
[35] CDF Collaboration, F. Abe *et al.*, Phys. Rev. **D41**, 2330 (1990).
[36] R. Engel, *Hadronic Interactions of Photons at High Energies*, 1997. Siegen (PhD Thesis).
[37] A. Moraes, C. Buttar, and I. Dawson, p. 46 (2005). ATL-PHYS-PUB-2005-007.
[38] J. M. Butterworth and S. Butterworth, Comput. Phys. Commun. **153**, 164 (2003). See results available at http://www.cedar.ac.uk/jetweb, hep-ph/0210404.
[39] F. Abe *et al.*, Phys. Rev. **D41**, 2330 (1990).
[40] G. Gustafson and G. Miu, Phys. Rev. **D63**, 034004 (2001). hep-ph/0002278.
[41] S. Catani, F. Fiorani, and G. Marchesini, Phys. Lett. **B234**, 339 (1990).
[42] M. Ciafaloni, Nucl. Phys. **B296**, 49 (1988).
[43] B. Andersson, G. Gustafson, and J. Samuelsson, Nucl. Phys. **B467**, 443 (1996).
[44] G. Gustafson, L. Lönnblad, and G. Miu, Phys. Rev. **D67**, 034020 (2003). hep-ph/0209186.
[45] A. H. Mueller, Nucl. Phys. **B415**, 373 (1994).
[46] A. H. Mueller and B. Patel, Nucl. Phys. **B425**, 471 (1994). hep-ph/9403256.
[47] A. H. Mueller, Nucl. Phys. **B437**, 107 (1995). hep-ph/9408245.
[48] E. Avsar, G. Gustafson, and L. Lönnblad, JHEP **07**, 062 (2005). hep-ph/0503181.
[49] V. A. Abramovsky, V. N. Gribov, and O. V. Kancheli, Yad. Fiz. **18**, 595 (1973).
[50] G. Gustafson, Z. Phys. **C15**, 155 (1982).
[51] J. M. Butterworth and M. Wing, Rep. Prog. Phys. **68**, 2773 (2005). hep-ex/0509018.