

# Tuning PYTHIA 8.1: the Monash 2013 tune

P. Skands<sup>1,a</sup>, S. Carrazza<sup>2</sup>, J. Rojo<sup>1,3</sup>

<sup>1</sup> Theoretical Physics, CERN, 1211 Geneva 23, Switzerland

<sup>2</sup> Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano, Via Celoria 16, 20133 Milan, Italy

<sup>3</sup> Rudolf Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, Oxford, UK

Received: 16 July 2014 / Accepted: 4 August 2014 / Published online: 15 August 2014  
© The Author(s) 2014. This article is published with open access at Springerlink.com

**Abstract** We present an updated set of parameters for the PYTHIA 8 event generator. We reevaluate the constraints imposed by LEP and SLD on hadronization, in particular with regard to heavy-quark fragmentation and strangeness production. For hadron collisions, we combine the updated fragmentation parameters with the new NNPDF2.3 LO PDF set. We use minimum-bias, Drell–Yan, and underlying-event data from the LHC to constrain the initial-state-radiation and multi-parton-interaction parameters, combined with data from SPS and the Tevatron to constrain the energy scaling. Several distributions show significant improvements with respect to the current defaults, for both  $ee$  and  $pp$  collisions, though we emphasize that interesting discrepancies remain in particular for strange particles and baryons. The updated parameters are available as an option starting from PYTHIA 8.185, by setting `Tune:ee = 7` and `Tune:pp = 14`.

## 1 Introduction

A truly impressive amount of results on QCD has been produced by the first run of the LHC. Most of these are already available publicly, e.g. via the data preservation site HEP-DATA [1]. A large fraction has also been encoded in the analysis preservation tool RIVET<sup>1</sup> [2]. Especially in the area of soft QCD, many of the experimental results have spurred further modelling efforts in the theory community (nice summaries of some of the current challenges can be found in [3,4]), while there is also significant activity dedicated to improving (“tuning”) the parameters of the existing models to better describe some or all of the available new data (see, e.g., the recent review in [5]).

<sup>1</sup> In particular, RIVET ensures that any (current or future) Monte Carlo event-generator codes can be compared consistently to the data, with exactly the same cuts, definitions, etc., as the original analysis.

<sup>a</sup> e-mail: peter.skands@cern.ch

The PYTHIA event generator [6,7] has been extensively compared to LHC data, and several tuning efforts have already incorporated data from Run 1 [5,8–16]. However, in particular for the newest version of the model, PYTHIA 8 [7], it has been some time since the constraints imposed by  $ee$  colliders were revised (in 2009), and then only via an undocumented tuning effort (using the PROFESSOR tool [17]). One of the main aims of this paper is therefore first to take a critical look at the constraints arising from LEP, SLD, and other  $e^+e^-$  experiments, reoptimize the final-state radiation and hadronization parameters, and document our findings. We do this manually, rather than in an automated setup, in order to better explain the reasoning behind each parameter adjustment. This writeup is thus also intended to function as an aid to others wishing to explore the PYTHIA 8 parameter space.

We then consider the corresponding case for hadron colliders, and use the opportunity to try out a new PDF set, an LO fit produced by the NNPDF collaboration [18–20] which has recently been introduced in PYTHIA 8 (NLO and NNLO sets are also available, for people that want to check the impact of using LO vs (N)NLO PDFs in hard-scattering events). In a spirit similar to that of the so-called “Perugia tunes” of PYTHIA 6 [8,21], we choose the same value of  $\alpha_s(M_Z) = 0.1365$  for both initial- and final-state radiation. (Though we do regard this choice as somewhat arbitrary, it may facilitate matching applications [21].) Again, we adjust parameters manually and attempt to give brief explanations for each modification. We also choose the  $\alpha_s(M_Z)$  value for hard-scattering matrix elements to be the same as that in the PDFs, here  $\alpha_s(M_Z) = 0.13$ . (The difference between the value used for radiation and that used for hard-scattering MEs may be interpreted as an artifact of translations between the CMW and  $\overline{MS}$  schemes, see Sect. 3.3.)

Below, in Sect. 1.1, we begin by giving a brief general explanation of the plots and  $\chi^2$  values that are used throughout the paper. Next, in Sect. 2, we describe the

physics, parameters, and constraints governing fragmentation in hadronic  $Z$  decays (final-state radiation and string fragmentation). We turn to hadron colliders in Sect. 3 (PDFs, initial-state radiation, and multi-parton interactions). We then focus on the energy scaling between different  $ee$  and  $pp$  ( $p\bar{p}$ ) collider energies in Sect. 4, including in particular the recently published high-statistics data from the Tevatron energy scan from 300 to 1960 GeV [22,23]. We round off with conclusions and a summary of recommendations for future efforts in Sect. 5.

A complete listing of the Monash 2013 tune parameters is given in Appendix A. Appendix B contains a few sets of additional plots, complementing those presented in the main body of the paper.

### 1.1 Plot legends and $\chi^2$ values

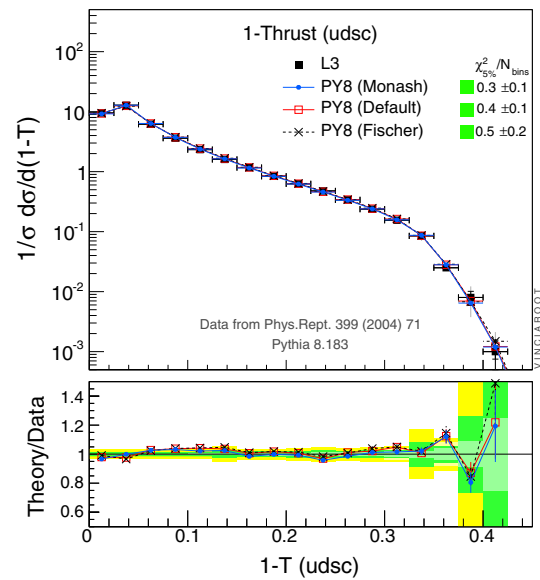
In several places, we have chosen to use data sets/constraints that differ from the standard ones available e.g. through RIVET (as documented below). Since our tuning setup is furthermore manual, rather than automated, we have in fact not relied on RIVET in this work (though we have made extensive use of HEPDATA [1]). Instead, we use the VINCIAROOT plotting tool [24], which we have here upgraded to include a simple  $\chi^2$  calculation, the result of which is shown on each plot.

Note that we include a blanket 5% “theory uncertainty” in the definition of the  $\chi^2$  value, representing a baseline sanity limit for the achievable accuracy of the modeling<sup>2</sup> that also gives a basic protection against overfitting. Note also that, rather than letting the MC uncertainty enter in the definition of the  $\chi^2$  value (and thereby risking that low statistics generate artificially low  $\chi^2$  values), we use the generated MC statistics to compute a  $\pm$  uncertainty on the calculated  $\chi^2$  value, which is also shown on the plots. Our definition of  $\chi^2$  is thus:

$$\left\langle \chi_{5\%}^2 \right\rangle = \frac{1}{N_{\text{bins}}} \sum_{i=1}^{N_{\text{bins}}} \frac{(\text{MC}_i - \text{Data}_i)^2}{\sigma_{\text{Data},i}^2 + (0.05\text{MC}_i)^2}, \quad (1)$$

with the corresponding MC uncertainty,  $\sigma_{\text{MC},i}$ , used to compute the statistical uncertainty on the  $\chi^2$  computation, as mentioned above. As is shown here, the normalization is always  $1/N_{\text{bins}}$ , regardless of whether the distributions are normalized to a fixed number or not, and we do not attempt to take into account correlations between the different observables. Since our tuning is not directly driven by a  $\chi^2$  minimization, we regard this as acceptable; the  $\chi_{5\%}^2$  values are intended merely to give an overall indication of the level of agreement or disagreement for each observable.

<sup>2</sup> We note that a similar convention is used on the MC PLOTS validation web site [25].



**Fig. 1** Hadronic  $Z$  decays at  $\sqrt{s} = 91.2$  GeV. The Thrust distribution in light-flavour tagged events, compared with L3 data [26]

The resulting plots look as illustrated in Fig. 1, with a main pane (top) showing the distribution itself and a bottom pane showing ratios. In the top pane, experimental data is always shown with filled black square symbols, with vertical black lines indicating the one-sigma uncertainties (with two separate black crossbars if separate statistical and systematic uncertainties are given). Lighter (grey) extensions of the vertical lines are used to indicate two-sigma uncertainties. In the ratio pane, the green shaded region indicates the one-sigma uncertainty region, while yellow is used to denote the two-sigma one. An internal lighter/darker shading variation in each band is used to denote the breakdown into statistical-only (inner) and statistical+systematic uncertainties (outer), whenever separate values for each of these are given. Finally, next to each MC legend the  $\chi_{5\%}^2$  value defined above is printed, along with its MC uncertainty. A colour-coded box next to the  $\chi^2$  value is shaded green ( $\chi^2 < 1$ ), yellow ( $1 < \chi^2 < 4$ ), orange ( $4 < \chi^2 < 9$ ), or red ( $9 < \chi^2$ ), depending on the level of agreement or disagreement. This functionality will be included in a forthcoming update of the VINCIAROOT plug-in to PYTHIA 8.

## 2 Final-state radiation and hadronization

The main parameter governing final-state radiation is the effective value of the strong coupling, which in PYTHIA 8 is specified by giving the value of  $\alpha_s(M_Z)$ . We follow the strategy of [24] and use a set of light-flavour ( $udsc$ ) tagged  $e^+e^-$  event shapes provided by the L3 experiment [26] to extract a best-fit value for  $\alpha_s(M_Z)$ . (This prevents  $B$  decays from contaminating this step of the analysis. Heavy-quark

fragmentation will be treated separately, below.) The renormalization scale for final-state shower emissions in PYTHIA is fixed to be [27]:

$$\text{FSR: } \mu_R^2 = p_{\perp\text{evol}}^2 = z(1-z)Q^2, \quad (2)$$

with  $Q^2 = p^2 - m_0^2$  the offshellness of the emitting parton (with on-shell mass  $m_0$ ), and  $z$  the energy fraction appearing in the DGLAP splitting kernels,  $P(z)$ . (To estimate the shower uncertainties associated with this choice of renormalization scale, we recommend using  $\ln(\mu_R^2) \pm \ln(2)$ , corresponding to a factor  $\sqrt{2}$  variation of  $\mu_R$ .)

Theoretically, a set of formally subleading terms can be resummed by using 2-loop running of  $\alpha_s$  in the so-called MC (a.k.a. CMW) scheme [28]. However, in a leading-order code like PYTHIA, this produces too little hard radiation in practice, due to missing NLO “K” factors for hard emissions (see, e.g., the study of NLO corrections in [29]). Empirically, we find that a better overall description is achieved with one-loop running, which, for a fixed value of  $\Lambda_{\text{QCD}}$ , can effectively mimic the effect of missing  $K$  factors via its relatively slower pace of running, leading to values of  $\alpha_s(M_Z)$  in the range 0.135–0.140, consistent with other LO extractions of the same quantity. (See [29] for an equivalent extraction at NLO.)

For this study, we did not find any significant advantage in reinterpreting this value in the CMW scheme<sup>3</sup> and hence merely settled on an effective  $\alpha_s(M_Z) = 0.1365$  (to be compared with the current default value of 0.1383).

For the infrared shower cutoff, we choose a value close to<sup>4</sup>  $\Lambda_{\text{QCD}}$ , in order to have a smooth transition between low- $p_{\perp}$  perturbative emissions and non-perturbative string breaks, the latter of which involve  $p_{\perp}$  kicks of order  $\Lambda_{\text{QCD}}$ . (In principle, the perturbative evolution could be continued to even lower scales, if combined with a non-perturbative regularization of  $\alpha_s$ , but such low cutoff values could risk generating problems at the fragmentation stage since the technical implementation of the string model becomes complicated if there are too many small gluon “kinks” spaced closely along the strings.) The set of relevant parameters in the code is:

```
# FSR: Strong Coupling
TimeShower:alphaSvalue = 0.1365
TimeShower:alphaSorder = 1
TimeShower:alphaSuseCMW = off
```

<sup>3</sup> One slight *disadvantage* is that the CMW scheme produces somewhat larger  $\Lambda_{\text{QCD}}$  values. Since the current formulation of the shower algorithm does not include a non-perturbative regularization of  $\alpha_s$ , a higher  $\Lambda_{\text{QCD}}$  value necessitates a larger IR cutoff in the shower, which can leave an undesirable gap between the transverse kicks generated by shower emissions and those generated by non-perturbative string splittings.

<sup>4</sup> The IR shower cutoff must still remain somewhat above the Landau pole of  $\alpha_s$ ; a lower cutoff scale would activate a hardcoded protection mechanism implemented in the PYTHIA shower, forcing it to be higher than  $\Lambda_{\text{QCD}}$ .

```
# FSR: IR cutoff
TimeShower:pTmin
= 0.50 ! for QCD radiation
TimeShower:pTminChgQ
= 0.50 ! for QED radiation off quarks
# FSR: Spin Correlations
TimeShower:phiPolAsym
= on ! approximate FSR polarization effects
```

The resulting distribution of the Thrust event-shape variable was shown in Fig. 1, comparing the Monash 2013 tune to the current default tune and to an alternative contemporary tune by Fischer [30]. To avoid clutter, the other event-shape variables ( $C$ ,  $D$ ,  $B_W$ , and  $B_T$ ) are collected in Appendix B.1. There are no significant changes to any of the light-flavour tagged event shapes in our tune as compared to the current default one.

## 2.1 Light-flavour fragmentation

Given a set of post-shower partons, resolved at a scale of  $Q_{\text{had}} \sim 1$  GeV, the non-perturbative stage of the fragmentation modeling now takes over, to convert the partonic state into a set of on-shell hadrons. In the leading-colour approximation, each perturbative dipole is dual to a non-perturbative string piece [31]. Quarks thus become string endpoints, while gluons become transverse kinks, connecting two string pieces [32]. The Lund string fragmentation model [33] describes the fragmentation of such string systems into on-shell hadrons.

Since the shower has already resolved all the (perturbative) physics down to a transverse-momentum scale of  $p_{T\text{min}} = 0.5$  GeV (for the Monash 2013 tune), we find it reasonable that the  $p_{\perp}$  kicks involved in string breaking should effectively average over dynamics in roughly the range  $250 \text{ MeV} = \sqrt{\kappa/\pi} < \sigma_{\perp} < p_{T\text{min}}$ , with the lower bound given by Fermi motion (with  $\kappa$  the string tension, see [34]). Further, since we here choose  $p_{T\text{min}}$  to be only slightly greater than  $\Lambda_{\text{QCD}}$ , the size of the non-perturbative corrections is naturally limited to kicks/corrections appropriate for non-perturbative dynamics (in contrast, e.g., to the cluster model [35], which can generate substantially larger kicks, of order the largest allowed cluster mass, which can be several GeV [30]). For the Monash 2013 tune, we have settled on a value of  $\sigma_{\perp} = 0.335$  GeV, with a small (1 %) tail of breaks involving higher  $p_{\perp}$  values carried over from the default settings.

```
StringPT:sigma = 0.335
StringPT:enhancedFraction = 0.01
StringPT:enhancedWidth = 2.0
```

This value is obtained essentially from the first two bins of the Thrust distribution, Fig. 1, and from the bins near zero of the other event shapes, see Appendix B.1. Note that the  $\sigma_{\perp}$  value is interpreted as the width of a Gaussian distribution in the total  $p_{\perp}$  (measured transversely to the local string

direction, which may differ from the global event axis), such that each of the  $p_x$  and  $p_y$  components have a slightly smaller average value,  $\sigma_{x,y}^2 = \frac{1}{2}\sigma_{\perp}^2 = (0.237 \text{ GeV})^2$ . Also note that each non-leading hadron will receive two  $p_{\perp}$  kicks, one from each of the breaks surrounding it, hence  $\langle p_{\perp, \text{had}}^2 \rangle = 2\sigma_{\perp}^2 = (0.474 \text{ GeV})^2$ .

For massless quarks, the longitudinal component of the energy carried by a hadron formed in the string-breaking process string  $\rightarrow$  hadron + string' is governed by the Lund symmetric fragmentation function:

$$f(z) \propto \frac{z^{(a_i - a_j)}(1 - z)^{a_j}}{z} \exp\left(\frac{-bm_{\perp}^2}{z}\right), \quad (3)$$

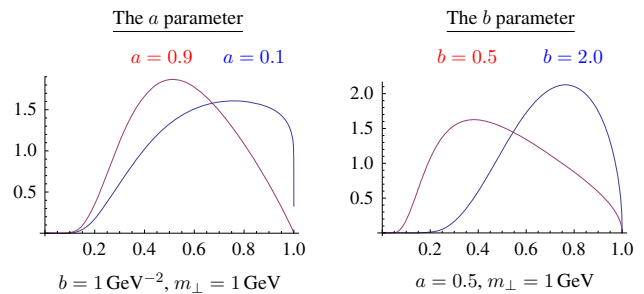
where  $z$  is the energy carried by the newly formed ( $ij$ ) hadron, expressed as a fraction of the (lightcone) energy of the quark (or antiquark) endpoint,  $i$ , of the fragmenting string. (The remaining energy fraction,  $(1 - z)$ , goes to the new string' system, from which another hadron can be split off in the same manner, etc., until all the energy is used up.) The transverse mass of the produced ( $ij$ ) hadron is defined by  $m_{\perp}^2 = m_{\text{had}}^2 + p_{\perp, \text{had}}^2$ , hence heavier hadrons have harder spectra. The proportionality sign in Eq. (3) indicates that the function is to be normalized to unity.

The  $a$  and  $b$  parameters govern the shape of the fragmentation function, and must be constrained by fits to data. Eq. (3) expresses the most general form of the fragmentation function, for which the  $a$  parameters of the original string-endpoint quark,  $a_i$ , and that of the (anti-)quark produced in the string break,  $a_j$ , can in principle be different, while the  $b$  parameter is universal. Within the Lund model, the  $a$  value is normally also taken to be universal, the same for all quarks, with the only freedom being that a larger  $a$  parameter can be assigned to diquarks [36], from which baryons are formed, and hence meson and baryon spectra can be decoupled somewhat. (See `StringZ:aExtraDiquark` below.)

Roughly speaking, large  $a$  parameters suppress the hard region  $z \rightarrow 1$ , while a large  $b$  parameter suppresses the soft region  $z \rightarrow 0$ . By adjusting them independently, both the average hardness and the width of the resulting fragmentation spectra can be modified. For example, increasing both  $a$  and  $b$  yields a narrower distribution, while changing them in opposite directions moves the average.

An illustration of the effect of varying the  $a$  and  $b$  parameters, for  $a_i = a_j \equiv a$ , is given in Fig. 2; see also the lecture notes in [37]. Note that the  $\sigma_{\perp}$  parameter also affects the hardness, with larger  $\sigma_{\perp}$  values generating harder hadrons, the difference being that the  $\sigma_{\perp}$  parameter acts mainly in the direction transverse to the string<sup>5</sup> (and is an absolute scale

<sup>5</sup> Explicitly,  $\sigma_{\perp}$  expresses the  $p_{\perp}$  broadening transverse to the string direction, but implicitly its size also enters in the longitudinal fragmentation function, via the  $m_{\perp}^2$  term in Eq. (3), causing higher- $p_{\perp}$  hadrons to have relatively harder longitudinal spectra as well.



**Fig. 2** Illustration of the Lund symmetric fragmentation function (normalized to unity), for  $a_i = a_j \equiv a$ . *Left* variation of the  $a$  parameter, from 0.1 (blue) to 0.9 (red), with fixed  $b$ . *Right* variation of the  $b$  parameter, from 0.5 (red) to 2 (blue)  $\text{GeV}^{-2}$ , with fixed  $a$

expressed in  $\text{GeV}$ ), while the  $a$  and  $b$  parameters act longitudinally (with  $z$  a relative scale expressed as a fraction of the endpoint's energy).

In the context of this work, we included the possibility of letting the  $a$  parameter for strange quarks be slightly different from that of  $u$  and  $d$  quarks, but did not find any significant advantages. The relevant parameters in the code we settled on for the Monash tune are:

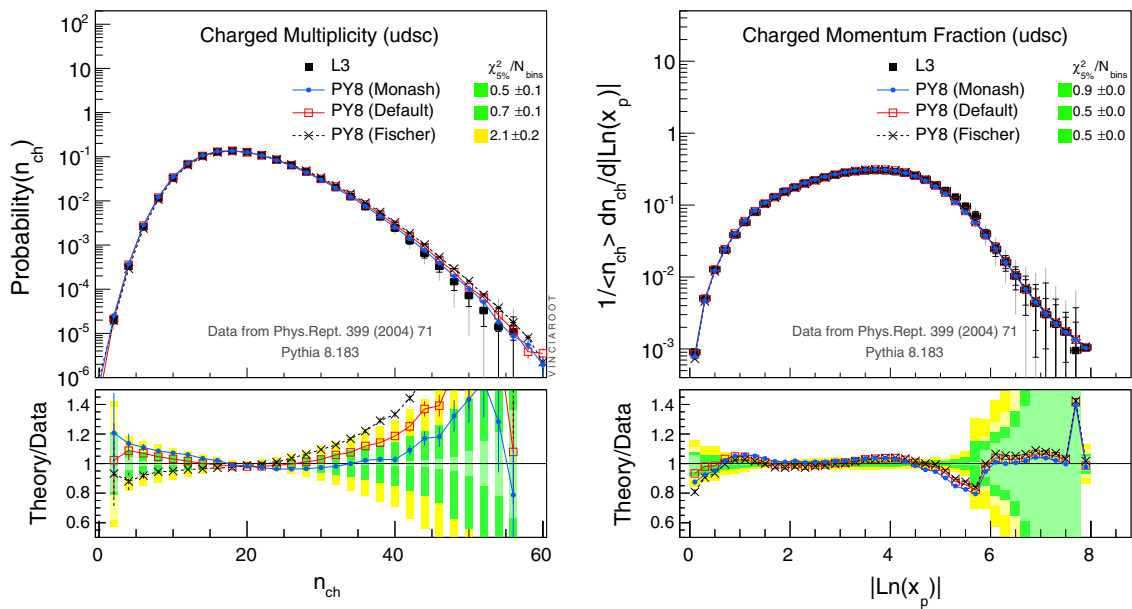
```
StringZ:aLund           = 0.68
StringZ:bLund           = 0.98
StringZ:aExtraDiquark  = 0.97
StringZ:aExtraSquark   = 0.00
```

The average hardness of the produced hadrons is tightly (anti-)correlated with the average multiplicity, via momentum conservation: if each hadron takes a lot of energy, then fewer hadrons must be made, and vice versa. Thus, the  $\sigma_{\perp}$  value and the  $a$  and  $b$  parameters of the fragmentation function can be well constrained by simultaneously considering both momentum and multiplicity spectra. In order to be as universal as possible, one normally uses the inclusive charged-particle spectra for this purpose. These are shown in Fig. 3. (Note: the Fischer tune only included the average particle multiplicity as a constraint, so the full  $n_{\text{ch}}$  distribution is not expected to be reproduced perfectly [30].) The momentum fraction in the right-hand plot is defined by:

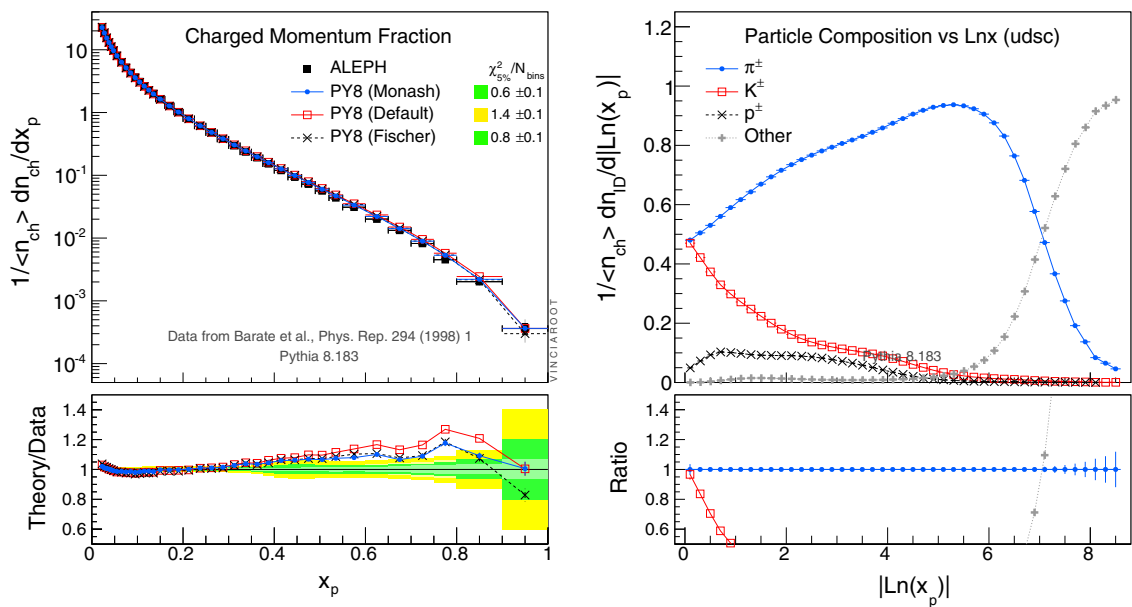
$$x_p = \frac{2|p|}{E_{\text{cm}}}. \quad (4)$$

As above, the experimental data come from a measurement by L3 [26] which only includes the four lightest flavours, thus excluding  $b$  quarks (which will be treated separately below).

Both of the earlier tunes exhibit a somewhat too broad multiplicity distribution in comparison with the L3 data. The relatively large Lund  $a$  and  $b$  values used for the Monash tune, combined with its large  $\sigma_{\perp}$  value, produce a narrower  $n_{\text{Ch}}$  spectrum, with in particular a smaller tail towards large multiplicities. All the tunes produce a sensible momentum spectrum. The dip around  $|\ln(x)| \sim 5.5$  corresponds to the



**Fig. 3** Hadronic Z decays at  $\sqrt{s} = 91.2$  GeV. Charged-particle multiplicity (left) and momentum-fraction (right) spectra



**Fig. 4** Hadronic Z decays at  $\sqrt{s} = 91.2$  GeV. Charged-particle momentum fraction  $x_p$ , on a linear scale (left) and relative particle composition (right) for the log-scale distribution shown in Fig. 3

extreme soft-pion tail, with momenta at or below  $\Lambda_{\text{QCD}}$ . We did not find it possible to remove it by retuning, since a smaller  $b$  parameter would generate significantly too high particle multiplicities and a smaller  $\sigma_{\perp}$  would lead to conflict with the event-shape distributions.

A zoom on the high-momentum tail is provided by the left-hand plot in Fig. 4, which shows a comparison on a linear momentum scale, to a measurement by ALEPH [38] (now including  $Z \rightarrow b\bar{b}$  events as well as light-flavour ones). All the tunes exhibit a mild overshooting of the data in the region

$0.5 < x_p < 0.8$ , corresponding to  $0.15 < |\ln(x)| < 0.7$ , in which no similar excess was present in the L3 comparison. We therefore do not regard this as a significant issue<sup>6</sup> but

<sup>6</sup> One might worry whether the effect could be due solely to the  $Z \rightarrow b\bar{b}$  events which are only present in the ALEPH measurement, and if so, whether this could indicate a significant mismodeling of the momentum distribution in  $b$  events. However, as we show below in the section on  $b$  fragmentation, the charged-particle momentum distribution in  $b$ -tagged events shows no excess in that region (in fact, it shows an undershooting).

note that the excess is somewhat milder in the Fischer and Monash tunes.

Further information to elucidate the structure of the momentum distribution is provided by the plot in the right-hand pane of Fig. 4, which uses the same  $|\ln(x)|$  axis as the right-hand plot in Fig. 3 and shows the relative particle composition in the Monash tune for each histogram bin. (The category “Other” contains electrons and muons from weak decays.) An interesting observation is that the relatively harder spectrum of Kaons implies that, for the highest-momentum bins, the charged tracks are made up of an almost exactly equal mixture of Kaons and pions, despite Kaons on average only making up about 10 % of the charged multiplicity.

## 2.2 Identified particles

Continuing on the topic of identified particles, we note that the extraction of the  $a$  and  $b$  parameters from the inclusive charged-particle distributions is made slightly more complicated by the fact that not all observed particles are “primary” (originating directly from string breaks); many lower-mass particles are “secondaries”, produced by prompt decays of more massive states (e.g.,  $\rho \rightarrow \pi\pi$ ), whose relative rates and decay kinematics therefore influence the spectra. In the  $e^+e^-$  measurements we include here, particles with  $c\tau < 100$  mm were treated as unstable, hence leading to secondaries. (For completeness, we note that the equivalent standard cut at the LHC is normally 10 mm.)

The particle composition in PYTHIA 8 was already tuned to a set of reference values provided by the PDG [39], and the default parameters do reasonably well, certainly for the most copiously produced sources of secondaries. Nonetheless, we have here reoptimized the flavour-selection parameters of the string-fragmentation model using a slightly different set of reference data, combining the PDG tables with information provided directly by the LEP experiments via HEPDATA [1]. Based on the level of agreement or disagreement between different measurements of the same particles, we have made our own judgement as to the level of uncertainty for a few of the particles, as follows. (Unless otherwise stated, we use the value from the PDG. Particles and antiparticles are implicitly summed over, and secondaries from particles with  $c\tau < 100$  mm are included.)

- The various LEP and SLD measurements of the  $\phi$  meson rate on HEPDATA are barely compatible. E.g., OPAL [40] reports  $\langle n_\phi \rangle = 0.091 \pm 0.002 \pm 0.003$  while ALEPH [38] quotes  $\langle n_\phi \rangle = 0.122 \pm 0.004 \pm 0.008$ , a difference of 30 % with uncertainties supposedly less than 10 %. DELPHI [41] and SLD [42] fall in between. The PDG value is  $\langle n_\phi \rangle = 0.0963 \pm 0.003$ , i.e., with a combined uncertainty

of just 3 %. We choose to inflate the systematic uncertainties and arrive at  $\langle n_\phi \rangle = 0.101 \pm 0.007$ .

- For  $\Lambda$  production, we use the most precise of the LEP measurements, by OPAL<sup>7</sup> [43],  $\langle n_\Lambda \rangle = 0.374 \pm 0.002 \pm 0.010$ , about 5 % lower than the corresponding PDG value.
- For  $\Sigma^\pm$  baryons, we use a combination of the two most recent LEP measurements, by L3 [44] for  $\Sigma^+ + \bar{\Sigma}^-$  and by DELPHI [45] for  $\Sigma^- + \bar{\Sigma}^+$ , for an estimated  $\langle n_{\Sigma^\pm} \rangle = 0.195 \pm 0.018$ , which is roughly 10 % higher than the PDG value.
- For  $\Sigma^0$  baryons, we use the most recent measurement, by L3 [44],  $\langle n_{\Sigma^0} \rangle = 0.095 \pm 0.015 \pm 0.013$ ; this is about 20 % larger than the PDG value. The L3 paper comments on their relatively high value by noting that L3 had the best coverage for low-momentum baryons, hence smaller model-dependent correction factors.
- For  $\Delta^{++}$  baryons, there are only two measurements in HEPDATA [46,47], which are mutually discrepant by about  $2\sigma$ . The DELPHI measurement is nominally the most precise, but OPAL gives a much more serious discussion of systematic uncertainties. We choose to increase the estimated extrapolation errors of the DELPHI measurement by 50 % and obtain a weighted average<sup>8</sup> of  $\langle n_{\Delta^{++}} \rangle = 0.09 \pm 0.017$ , 5 % larger than the PDG value, with a 20 % larger uncertainty.
- For  $\Sigma^*$ , the three measurements on HEPDATA [38,43,48] are likewise discrepant by  $2\sigma - 3\sigma$ . We inflate the systematic uncertainties and arrive at  $\langle n_{\Sigma^{*\pm}} \rangle = 0.050 \pm 0.006$ , which is again 5 % higher than the PDG value, with twice as much uncertainty.
- The measurements for  $\Xi^\pm$  are in good agreement [38,43,48], with a weighted average of  $\langle n_{\Xi^\pm} \rangle = 0.0266 \pm 0.0012$ , slightly larger than the PDG value.
- For  $\Xi^{*0}$ , however, the DELPHI measurement [48] gives a far lower number than the OPAL [43] and ALEPH [38] ones, and the weighted average differs by more than 10 % from the PDG value, despite the latter claiming an uncertainty smaller than 10 %. Our weighted average is  $\langle n_{\Xi^{*0}} \rangle = 0.0059 \pm 0.0012$ .
- Finally, for the  $\Omega$  baryon, the DELPHI [49] and OPAL [43] measurements are in agreement, and we use the PDG value,  $\langle n_\Omega \rangle = 0.0016 \pm 0.0003$ .

We summarize the constraints on the light-meson and baryon rates used here in Table 1. Note that we express them as

<sup>7</sup> We note that HEPDATA incorrectly gives the systematic error as 0.002 while the value in the OPAL paper is 0.010 [43]. This has been communicated to the HEPDATA maintainers.

<sup>8</sup> Even with the inflated error, the uncertainty on the DELPHI measurement is still less than a third that of the OPAL one. DELPHI therefore still dominates the average.

**Table 1** Hadronic  $Z$  decays at  $\sqrt{s} = 91.2$  GeV. Measured rates of light-flavour mesons and baryons, expressed as percentages of the average charged-particle multiplicity, as used in this work. Multiply the numbers by 20.7/100 to translate the percentages to corresponding production rates. Source labels indicate: A (ALEPH), D (DELPHI), L (L3), O (OPAL), S (SLD), P (PDG)

Mesons $\langle n \rangle / \langle n_{Ch} \rangle$	Our reference value (in %)	Our source	Baryons $\langle n \rangle / \langle n_{Ch} \rangle$	Our reference value (in %)	Our source
$\pi^+ + \pi^-$	$82.2 \pm 0.9$	P	$p + \bar{p}$	$5.07 \pm 0.16$	P
$\pi^0$	$45.5 \pm 1.5$	P	$\Lambda + \bar{\Lambda}$	$1.81 \pm 0.051$	O
$K^+ + K^-$	$10.8 \pm 0.3$	P	$\Sigma^+ + \Sigma^- + \bar{\Sigma}^+ + \bar{\Sigma}^-$	$0.942 \pm 0.087$	DL
$\eta$	$5.06 \pm 0.38$	P	$\Sigma^0 + \bar{\Sigma}^0$	$0.459 \pm 0.096$	L
$\eta'$	$0.73 \pm 0.09$	P	$\Delta^{++} + \bar{\Delta}^{--}$	$0.434 \pm 0.082$	DO
$\rho^+ + \rho^-$	$11.6 \pm 2.1$	P	$\Sigma^{*+} + \Sigma^{*-} + \bar{\Sigma}^{*+} + \bar{\Sigma}^{*-}$	$0.242 \pm 0.029$	ADO
$\rho^0$	$5.95 \pm 0.47$	P	$\Xi^+ + \bar{\Xi}^-$	$0.125 \pm 0.0050$	ADO
$K^{*+} + K^{*-}$	$3.45 \pm 0.28$	P	$\Xi^{*0} + \bar{\Xi}^{*0}$	$0.0285 \pm 0.0058$	ADO
$\omega$	$4.90 \pm 0.31$	P	$\Omega^- + \bar{\Omega}^+$	$0.0077 \pm 0.0015$	P
$\phi$	$0.49 \pm 0.035$	ADOS			

percentages of the average charged multiplicity,

$$\langle n_{Ch} \rangle = 20.7, \tag{5}$$

obtained as a weighted average over MARK-II [50], ALEPH [38], DELPHI [51], OPAL [52], and L3 [53] measurements.

The light-flavour-selection parameters for the Monash tune are (see Appendix A for a comparison of these values to the current default ones):

```
# Light-Meson Sector
StringFlav:ProbStoUD      = 0.217
StringFlav:mesonUDvector  = 0.5
StringFlav:mesonSvector   = 0.55
StringFlav:etaSup         = 0.60
StringFlav:etaPrimeSup    = 0.12
# Baryon Sector
StringFlav:probQQtoQ      = 0.081
StringFlav:probSQtoQQ     = 0.915
StringFlav:probQQ1toQQ0  = 0.0275
StringFlav:suppressLeadingB = off
StringFlav:popcornSpair   = 0.9
StringFlav:popcornSmeson  = 0.5
```

Since strange-particle and baryon spectra at the LHC exhibit interesting differences with respect to existing models (see below), we paid particular attention to first obtaining a good description of these sectors in  $e^+e^-$  collisions. Specifically, we have increased the overall amount of strangeness by about 10 %, while decreasing the rate of vector mesons by a similar amount<sup>9</sup> (these two effects largely cancel for  $K^*$ ). This improves the total  $K^\pm$ ,  $\rho^0$ ,  $\omega$ ,  $\Lambda$ ,  $\Xi^*$ , and  $\Omega$  yields on our combined LEP estimates discussed above. The price is that we now overshoot the measured rate of  $\Xi^\pm$  baryons

<sup>9</sup> For reference, the current default value of ProbStoUD is 0.19 while ours is 0.217. The increased value also improves the agreement with the  $D_s$  and  $B_s$  rates, see Sect. 2.3. The default values of mesonUDvector and mesonSvector are 0.62 and 0.725 respectively, while ours are 0.5 and 0.55.

by 10 %. The resulting identified-meson and -baryon rates, expressed as fractions of the average charged-particle multiplicity are plotted in Fig. 5. Note that the last four bins of the meson plot and the third and fourth bins of the baryon plot are not  $\langle n \rangle / \langle n_{Ch} \rangle$  fractions, but rather the  $K^*/K$ ,  $\phi/K^*$ ,  $\phi/K$ ,  $\phi/\pi$ ,  $\Lambda/p$  and  $\Lambda/K$  ratios, respectively. Note also that Sect. 4 on energy scaling below includes a comparison to the average Kaon and Lambda rates as a function of  $ee$  CM energy (Fig. 25).

To provide further information on identified particles, we include a limited comparison to momentum spectra of  $K^\pm$ ,  $p$ ,  $\Lambda$ , and  $\Xi^\pm$ , which are the states of most immediate interest in the context of similar comparisons now being made at LHC. The spectra of  $K^\pm$  mesons and  $\Lambda$  baryons are shown in Fig. 6, while the  $p^\pm$  and  $\Xi^\pm$  spectra are relegated to Appendix B.2. The modified parameters of the Monash tune have virtually no effect on the Kaon distribution, which still exhibits too many very soft Kaons (with  $\ln(x) < -4$ , corresponding to  $x < 0.018$ , so momentum scales below  $\sim 1$  GeV), while the significant increase in the value of aExtraDiquark from 0.5 (Default) to 0.97 (Monash, cf. Sect. 2.1) produces a desirable suppression of very hard  $\Lambda$  baryons. The corresponding change in the measured parts of the  $p$  and  $\Xi^\pm$  spectra (cf. Appendix B.2) are small compared with the experimental uncertainties.

It is interesting, however, to note that all of these spectra indicate, or are at least consistent with, a modelling excess of soft identified-particle production below  $\ln(x) \sim -4.5$ , corresponding to absolute momentum scales around 500 MeV, while we recall that the inclusive  $\ln(x)$  spectrum above showed an underproduction around  $\ln(x) \sim -5.5$ . Within the constraints of the current theory model, we have not managed to find a way to mitigate these features while remaining consistent with the rest of the data. Nonetheless, it should be mentioned that these observations could have relevance also

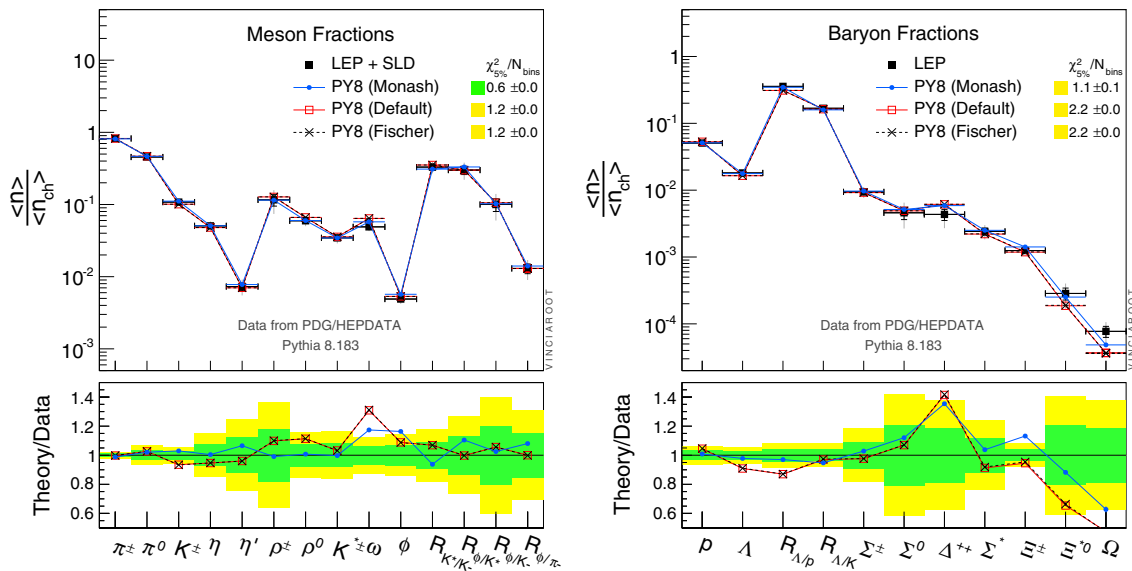


Fig. 5 Hadronic Z decays at  $\sqrt{s} = 91.2$  GeV. Identified-meson and -baryon rates, expressed as fractions of the average charged-particle multiplicity

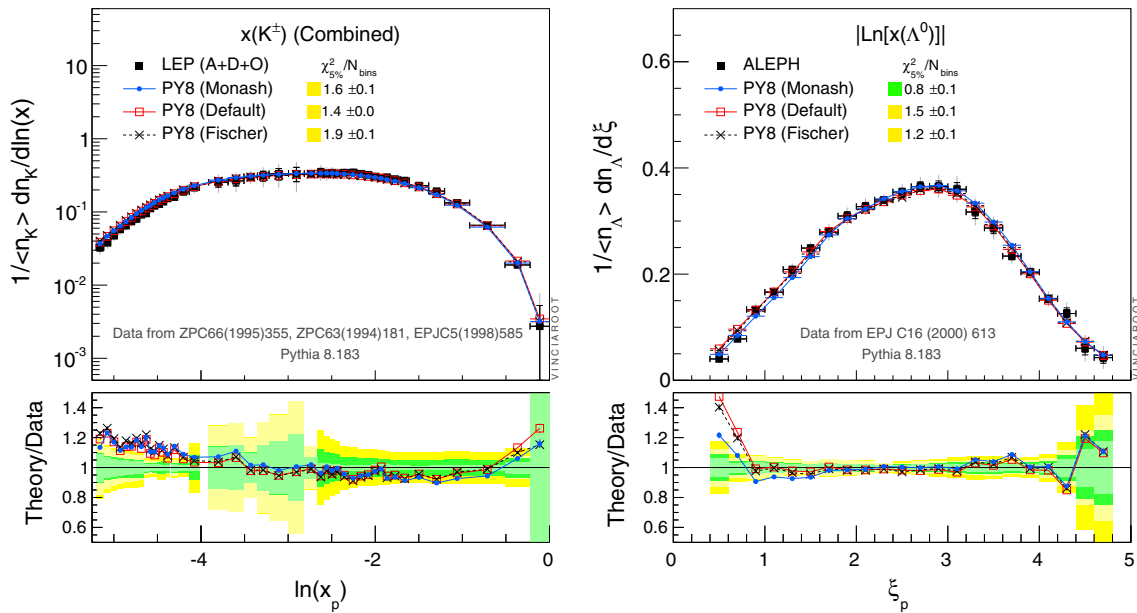


Fig. 6 Hadronic Z decays at  $\sqrt{s} = 91.2$  GeV.  $K^\pm$  and  $\Lambda$  momentum-fraction spectra

in the context of understanding identified-particle spectra at LHC, a possibility which to our knowledge has so far been ignored.

### 2.3 Heavy-quark fragmentation

Similarly to above, we first discuss the inclusive rates of hadrons containing heavy quarks, before we discuss their spectra. Unfortunately, there are also here substantial disagreements between different pieces of information. We have made the following choices.

- For  $D$  mesons, the average  $D^\pm$  rate given in sec. 46 of the PDG (0.175) is equal to the inclusive branching fraction for  $Z \rightarrow D^\pm X$  given in the  $Z$  boson summary table in the same *Review* (after normalizing the latter to the hadronic  $Z$  fraction of 69.91 % [39]). However, the former ought to be substantially larger given that some  $Z \rightarrow c\bar{c}$  events will contain two  $D^\pm$  mesons (counting once in the  $Z \rightarrow D^\pm X$  branching fraction but twice in the average  $D^\pm$  multiplicity). We therefore here use a measurement by ALEPH [54] to fix the  $D^\pm$  and  $D^0$  rates, resulting in a reference value for the average  $D^\pm$  multiplicity almost twice as large as that given by sec. 46 in the PDG.



**Table 2** Hadronic  $Z$  decays at  $\sqrt{s} = M_Z$ . Measured rates and inclusive branching fractions of particles containing  $c$  and  $b$  quarks, as used in this work. Note The branching fractions are normalized to  $Z \rightarrow$  hadrons, and hence should be interpreted as, e.g.,  $\text{BR}(Z \rightarrow B^+X)/\text{BR}(Z \rightarrow$

hadrons). Note 2 The sum over  $B^*$  states includes both particles and anti-particles. Note 3 The  $\Upsilon$  rate is multiplied by a factor 10. Source labels indicate: A (ALEPH), D (DELPHI), O (OPAL), P (PDG, section 46), S (SLD), Z (PDG Z Boson Summary Table)

Charm $\langle n \rangle$ or BR	Our reference value	Our source	Beauty $\langle n \rangle$ or BR	Our reference value	Our source
$D^+ + D^-$	$0.251 \pm 0.047$	A	$\text{BR}(Z \rightarrow B^+X)$	$0.087 \pm 0.002$	Z
$D^0 + \bar{D}^0$	$0.518 \pm 0.063$	A	$B^+ + B^0 + \bar{B}^0 + B^-$	$0.330 \pm 0.052$	P
$D^{*+} + D^{*-}$	$0.194 \pm 0.0057$	P	$B_u^* + B_d^* + B_s^*$	$0.288 \pm 0.026$	P
$D_s^+ + D_s^-$	$0.131 \pm 0.021$	P	$\text{BR}(Z \rightarrow B_s^0X)$	$0.0227 \pm 0.0019$	Z
$\text{BR}(Z \rightarrow \Lambda_c^+X)$	$0.0220 \pm 0.0047$	Z	$\text{BR}(Z \rightarrow B_{\text{baryon}}X)$	$0.0197 \pm 0.0032$	Z
$\text{BR}(Z \rightarrow X + c\bar{c})$	$0.0306 \pm 0.0047$	AO	$\text{BR}(Z \rightarrow X + b\bar{b})$	$0.00288 \pm 0.00061$	ADS
$J/\psi$	$0.0052 \pm 0.0004$	P	$\text{BR}(Z \rightarrow b\bar{b}b\bar{b}X)$	$0.00051 \pm 0.00019$	Z
$\chi_{c1}$	$0.0041 \pm 0.0011$	P	$\Upsilon (\times 10)$	$0.0014 \pm 0.0007$	P
$\psi'$	$0.0023 \pm 0.0004$	P			

- For  $\Lambda_c^+$ , the average multiplicity given in sec. 46 of the PDG is twice as large as that indicated by the branching fraction  $\text{BR}(Z \rightarrow \Lambda_c^+X)$  in the  $Z$  boson summary table in the same *Review*. We here use the branching from the  $Z$  boson summary table as our constraint on the  $\Lambda_c^+$  rate, normalized to the total branching fraction  $\text{BR}(Z \rightarrow \text{hadrons})$ .
- We also include the average rate of  $g \rightarrow c\bar{c}$  splittings, obtained by combining an ALEPH [55] and an OPAL measurement [56], but with an additional 10 % systematic uncertainty added to both measurements to account for possibly larger mismodeling effects in the correction factors [57,58].
- For  $B$  particles, we use the quite precise inclusive  $Z \rightarrow B^+X$  branching fraction from the  $Z$  boson summary in the PDG.
- We also use the sum of  $B^\pm$  and  $B^0(\bar{B}^0)$  in sec. 46 of the PDG.<sup>10</sup>
- The  $B_s^0$  multiplicity given in sec. 46 of the PDG (0.057  $\pm$  0.013) is more than twice the inclusive  $\text{BR}(Z \rightarrow B_s^0X)/\text{BR}(Z \rightarrow \text{hadrons})$  branching fraction (0.0227  $\pm$  0.0019) quoted in the  $Z$  boson summary table. We find these two numbers difficult to reconcile and choose to use the inclusive  $\text{BR}(Z \rightarrow B_s^0X)/\text{BR}(Z \rightarrow \text{hadrons})$  branching fraction as our main constraint.
- We also include the inclusive branching fractions for  $B$ -baryons (summed over baryons and antibaryons), the rate of  $g \rightarrow b\bar{b}$  splittings obtained by combining ALEPH [59], DELPHI [60], and SLD [61] measurements (including an additional 10 % systematic to account for larger possible

mismodeling effects in the correction factors [57,58]) and the rate of  $Z \rightarrow b\bar{b}b\bar{b}$  from the PDG  $Z$  boson summary table [39].

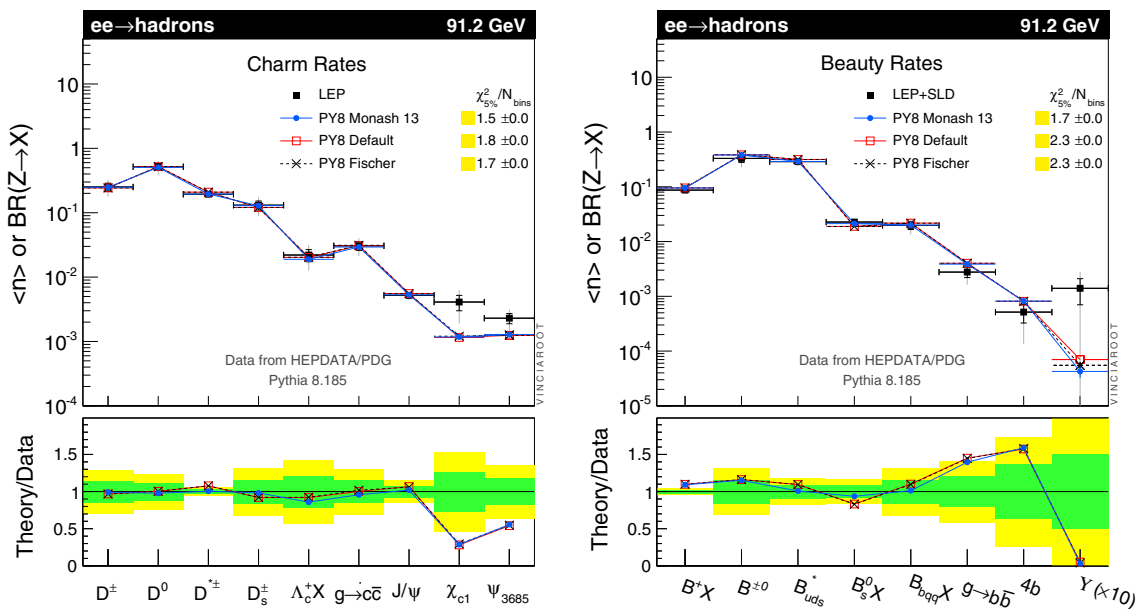
Our constraints on the heavy-quark particle rates are summarized in Table 2. Comparisons to these rates are shown in Fig. 7, now without normalizing to the average charged-particle multiplicity. Given that most of the  $c$  and  $b$  quarks come directly from  $Z \rightarrow c\bar{c}$  and  $Z \rightarrow b\bar{b}$  decays, there is not a lot of room for tuning to these numbers, apart from the relative rates of vector mesons vs. pseudoscalars, which is controlled by the parameters:

```
# Heavy Mesons
StringFlav:mesonCvector = 0.88
StringFlav:mesonBvector = 2.2
```

Our parameters are slightly smaller than the current default values, leading to slightly smaller  $D^*$  and  $B^*$  rates, as can be seen from the plots in Fig. 7. Note also that the increased overall amount of strangeness in the fragmentation leads to slightly higher  $D_s$  and  $B_s$  fractions, in better agreement with the data. Uncertainties are, however, large, and some exotic onium states, like  $\chi_{c1}$ ,  $\psi'$ , and  $\Upsilon$  are not well described by the default modeling. (It is encouraging that at least the multiplicity of  $J/\psi$  mesons is well described, though a substantial fraction of this likely owes to the feed-down from  $B$  decays, and hence does not depend directly on the string-fragmentation model itself.)

We also note that it would be desirable to reduce the rate of  $g \rightarrow b\bar{b}$  and  $Z \rightarrow b\bar{b}b\bar{b}$  events, while the  $g \rightarrow c\bar{c}$  one appears consistent with the LEP constraints. We suspect that this issue may be tied to the fixed choice of using  $p_\perp$  as the renormalization scale for both gluon emissions and for  $g \rightarrow q\bar{q}$  splittings in the current version of PYTHIA. A more natural choice for  $g \rightarrow q\bar{q}$  could be  $\mu_R \propto m_{q\bar{q}}$ , as used e.g. in the VINCIA shower model [29].

<sup>10</sup> Note that we have a factor 2 relative to the PDG, since it appears the PDG quotes the average, rather than the sum. Note also that all the average  $B$  meson multiplicities in sec. 46 of the PDG are accompanied by a note, “(d)”, stating that the SM  $B(Z \rightarrow b\bar{b}) = 0.217$  was used for the normalization. For completeness, the reader should be aware that this is the fraction normalized to hadronic  $Z$  decays; the branching fraction relative to all  $Z$  decays, is 0.151 [39].



**Fig. 7** Hadronic  $Z$  decays at  $\sqrt{s} = 91.2$  GeV. Rates and inclusive  $Z \rightarrow X$  branching fractions (normalized to  $Z \rightarrow$  hadrons) of particles containing  $c$  and  $b$  quarks

We now turn to the dynamics of heavy-quark fragmentation, focusing mainly on the  $b$  quark.

For heavy quarks, the Lund fragmentation function is modified due to the (massive) endpoints not moving along straight lightcones: as the string pulls on them, they slow down, resulting in the string tracing out a smaller space-time area than it would for massless quarks. This modifies the implications of the string area law, in a manner captured by the so-called Bowler modification of the fragmentation function [62]

$$f_{\text{massive}}(z, m_Q) \propto \frac{f(z)}{z^{br_Q m_Q^2}}, \tag{6}$$

with  $m_Q$  the heavy-quark mass,  $b$  the same universal parameter that appears in the massless fragmentation function, Eq. (3), and  $r_Q$  a tuning parameter which is unity in the original derivation of Bowler but can be assigned values different from unity to reduce ( $r_Q \rightarrow 0$ ) or emphasize ( $r_Q > 1$ ) the effect. Since  $r_Q$  multiplies the heavy-quark mass (squared), it can also be viewed as an effective rescaling of the mass value. The net result is a suppression of the region  $z \rightarrow 1$ , hence a relative softening of the fragmentation spectrum for heavy flavours (relative since the presence of  $m_\perp^2$  in the exponent of Eq. (3) still implies an overall harder fragmentation for higher hadron masses.)

We emphasize that this is the only fragmentation function that is self-consistent within the string-fragmentation model [33,62]. Although a few alternative forms of the fragmentation functions for massive quarks are available in the code, we therefore here work only with the Bowler type. As for the

massless function, the proportionality sign in Eq. (6) indicates that the function is normalized to unity.

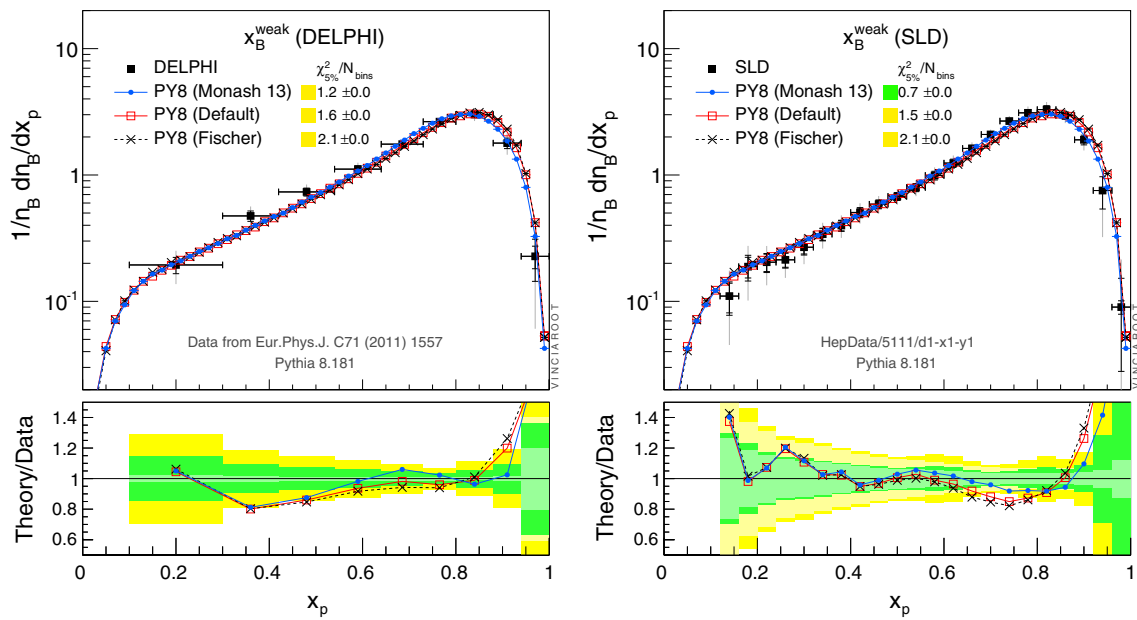
In PYTHIA, separate  $r_Q$  parameters are provided for  $c$  and  $b$  quarks. We consider the one for  $b$  quarks first. Its default value is  $r_b = 0.67$ , but this appears to give too hard  $b$  fragmentation spectra when compared to LEP and SLD data, see below. For the Monash tune, we instead use

$$\text{StringZ:rFactB} = 0.855$$

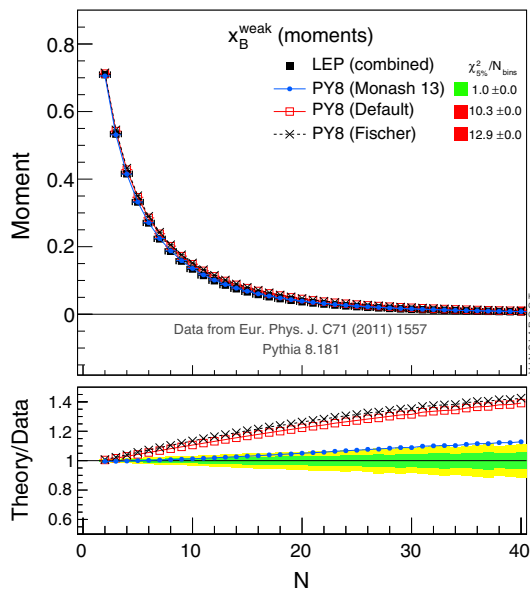
which produces softer  $B$  spectra and simultaneously agrees better with the theoretically preferred value ( $r_b = 1$ ).

A comparison to the scaled-momentum spectra ( $x_B = 2|p_B|/E_{\text{cm}}$ ) of weakly decaying  $B$  hadrons from both DELPHI [63] and SLD [64] is given in Fig. 8 (due to small differences between the two measured results, we choose to show both). The dampening of the hardest part of the spectrum caused by the increase in the  $r_b$  parameter is visible in the right-most two bins of the distributions and in the smaller  $\chi^2_{5\%}$  values for the Monash tune. The effects of the modification can be further emphasized by an analysis of the moments of the distribution, in which the higher moments are increasingly dominated by the region  $x_B \rightarrow 1$ . A comparison to a combined LEP analysis of the moments of the  $x_B$  distribution [63] is given in Fig. 9, further emphasizing that the high- $x_B$  part of the distribution is now under better control.

The reason we have not increased the  $r_b$  parameter further is that it comes at a price. If the  $B$  hadrons are taking less energy, then there is more energy left over to produce other particles, and the generated multiplicity distribution in  $b$  events already exhibits a slightly high tail towards large



**Fig. 8** Hadronic Z decays at  $\sqrt{s} = 91.2$  GeV. Momentum ( $x_B$ ) spectra of weakly decaying B hadrons, compared to data from DELPHI [63] (left) and SLD [64] (right)



**Fig. 9** Hadronic Z decays at  $\sqrt{s} = 91.2$  GeV. Moments of the B fragmentation function, compared to a combined analysis of LEP+SLD data by DELPHI [63]

multiplicities. Nonetheless, since the revised light-flavour fragmentation parameters produce an overall narrower fragmentation function, the end result is still a slight improvement in the multiplicity distribution also for  $b$  events. This is illustrated, together with the inclusive momentum distribution for  $b$ -tagged events, in Fig. 10, compared to measurements by L3 [26]. Interestingly, the multiplicity distribution still appears to be too wide, but within the constraints of the present study, we were unable to obtain further improvements. As a point

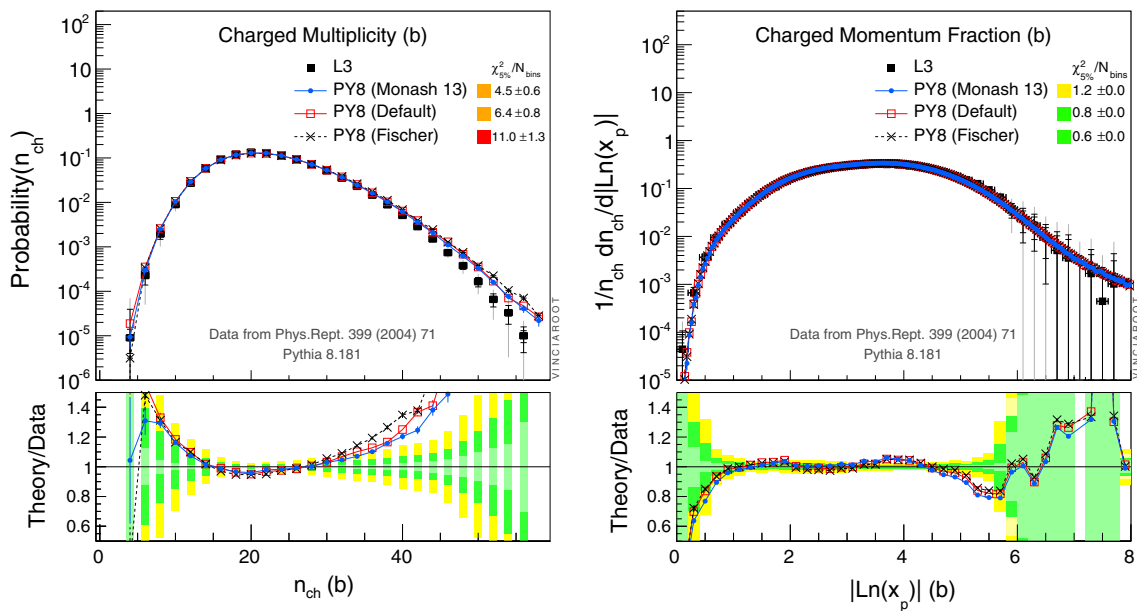
of speculation, we note that the distribution of the number of partons before hadronization is also quite wide in PYTHIA, and this may be playing a role in effectively setting a lower limit on the width that can be achieved for the hadron-level distribution.

Comparisons to L3 event shapes in  $b$ -tagged events are collected in Appendix B.1 (the left column of plots contains light-flavour tagged event shapes, the right column  $b$ -tagged ones). In particular, the Monash tune gives a significant improvement in the soft region of the jet-broadening parameters in  $b$ -tagged events, while no significant changes are observed for the other event shapes. These small improvements are presumably a direct consequence of the softening of the  $b$  fragmentation function; it is now less likely to find an isolated ultra-hard  $B$  hadron.

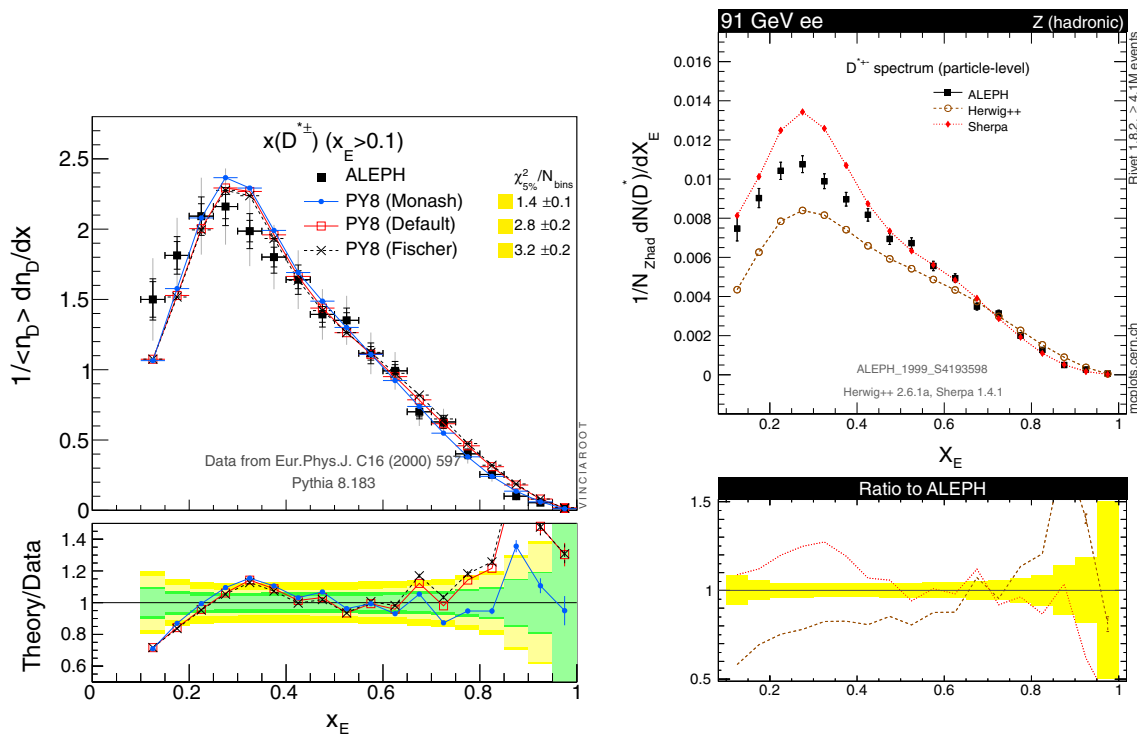
We round off the discussion of heavy-quark fragmentation by noting that a similarly comprehensive study of charm-quark fragmentation would be desirable. However, charm-quark tagged multiplicity and event-shape data is not available to our knowledge, and most of the  $D$  meson spectra on HEPDATA concern only specific decay chains (hence depend on the decay modeling), and/or are limited to restricted fiducial regions (limiting their generality). Experimentally, the cleanest measurement is obtained from  $D^*$  decays, and an inclusive momentum spectrum for  $D^*$  mesons has been measured by ALEPH [55]. From this distribution, shown in Fig. 11, we determine a value for  $r_c$  of:

$$\text{StringZ:rFactC} = 1.32$$

We note that the low- $x$  part of the  $D^*$  spectrum originates from  $g \rightarrow c\bar{c}$  shower splittings, while the high- $x$  tail repre-



**Fig. 10** Hadronic  $Z$  decays at  $\sqrt{s} = 91.2$  GeV. Charged-hadron multiplicity (*left*) and momentum-fraction (*right*) spectra in  $b$ -tagged events



**Fig. 11** The inclusive  $D^*$  spectrum in hadronic  $Z$  decays [55]. *Left* Monash 2013 tune compared with default PYTHIA 8 and the Fischer tune. *Right* Comparison with HERWIG (*dashed*) and SHERPA (*dot-*

*ted*), from MCPLOTS [25]. Note that the plot in the left-hand pane is normalized to unity, while the one in the *right-hand* pane is normalized to the number of hadronic  $Z$  decays

sents prompt  $D^*$  production from leading charm in  $Z \rightarrow c\bar{c}$  (see [55] for a nice figure illustrating this). The intermediate range contains a large component of feed-down from  $b \rightarrow c$  decays, hence this distribution is also indirectly sensitive to the  $b$ -quark sector. The previous default tune had a harder

spectrum for both  $b$ - and  $c$ -fragmentation, leading to an overestimate of the high- $x$  part of the  $D^*$  distribution. The under-shooting at low  $x_{D^*}$  values, which remains unchanged in the Monash tune, most likely indicates an underproduction of  $g \rightarrow c\bar{c}$  branchings in the shower. We note that such an

underproduction may also be reflected in the LHC data on  $D^*$  production, see e.g. [65]. We return to this issue in the discussion of identified particles at LHC, Sect. 3.5.

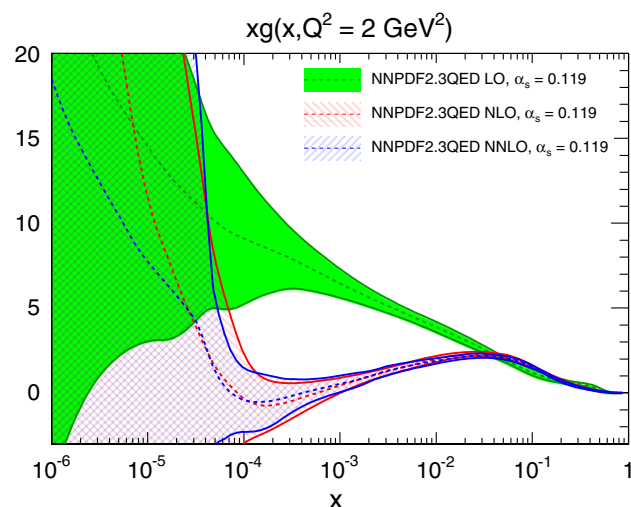
For completeness, the right-hand pane of Fig. 11 shows the  $D^*$  spectra from the two other general-purpose MC models, HERWIG [66] and SHERPA [67]. The HERWIG spectrum (dashed lines) is similar to the default PYTHIA one, with a deficit in the  $g \rightarrow c\bar{c}$  dominated region at low  $x_E$  and a significant overshooting in the hard leading-charm region,  $x_E \rightarrow 1$ . Interestingly, the  $D^*$  spectrum in SHERPA (dotted lines) exhibits an *excess* at small  $x_E$  values, suggesting relatively larger contributions from  $b$  decays and from  $g \rightarrow c\bar{c}$  splittings.

### 3 Hadron collisions

We discuss PDFs in Sect. 3.1, the choice of strong coupling (and total cross sections) in Sect. 3.2, initial-state radiation (and primordial  $k_T$ ) in Sect. 3.3, minimum-bias and underlying event in Sect. 3.4, and finally identified-particle spectra in Sect. 3.5. Energy scaling is discussed separately, in Sect. 4.

#### 3.1 Parton distributions

In the context of MC models, a highly important role is played by the small- $x$  gluon PDF, which has a strikingly different behavior between LO and NLO/NNLO fits. This effect is illustrated in Fig. 12, obtained from the NNPDF2.3QED PDF sets [19] (see also the useful plot of colour-weighted parton fluxes, fig. 2 in [13]). The origin of this different small- $x$  behavior is the missing large higher-order corrections to the DIS splitting functions and matrix elements (represented by



**Fig. 12** Comparison of the gluon PDF at  $Q^2 = 2 \text{ GeV}^2$  between the LO, NLO and NNLO fits of the NNPDF2.3QED family

coefficient functions) in the LO fit. Another source of the differences between LO and N(N)LO is related to the positivity of PDFs. Indeed, while at LO PDFs have a probabilistic interpretation and are thus positive-definite, starting from NLO they are scheme-dependent quantities and thus can become negative [68]. (Of course, physical observables like structure functions are positive-definite to all orders in the perturbative expansion.)

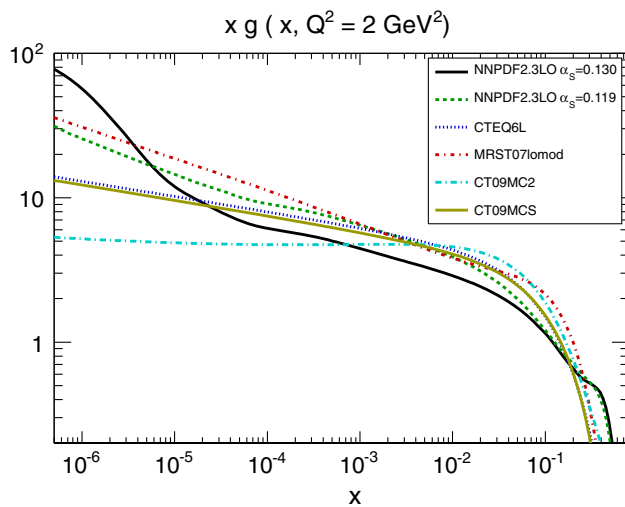
In recent years there has been some discussion about possible modifications of the vanilla LO PDFs that could lead to improved predictions from LO event generators. Some possibilities for these improvements that have been explored include the use of the LO value of  $\alpha_s$  but with two-loop running, or relaxing the momentum sum rules constraint from the LO fits. These and other related ideas underlie recent attempts to produce modified LO PDFs such as MRST2007lomod PDFs [69] and the CT09MC1/MC2 [70] PDFs. The claim was that such improved LO (also called LO\*) PDFs lead to a better agreement between data and theory in the LO fit and that their predictions for some important collider observables are closer to the results using the full NLO calculation. We note, however, that in the context of earlier multi-parton-interaction-model tuning studies undertaken by us [8] and by ATLAS [13], the large gluon component in LO\* PDFs has been problematic (driving very high inclusive-jet and MPI rates).

In the context of the NNPDF fits, which we shall use for the Monash 2013 tune, the above modifications were also studied. In particular, in the study of the NNPDF2.1LO fits in Ref. [18], it was found that, from the point of view of the agreement between data and theory, the standard LO PDFs provided as good a description as the other possible variations, including a different value of  $\alpha_s(M_Z)$ , using the one- or two-loop running or relaxing the momentum sum rule. The different results found by previous studies could be related to the limited flexibility in the input gluon PDFs in the CTEQ/MSRT LO fits: indeed, with a flexible enough parametrization such as that used in the NNPDF fits, the differences between these theory choices can always be absorbed into the initial condition.

Therefore, we have settled on an unmodified LO PDF set for the Monash 2013 tune, the NNPDF2.3 LO set [19,20], which combines the NNPDF2.1 LO PDFs with a determination of the photon PDF and a combined QCD+QED evolution [19,71]. The relevant parameter in the code is:

```
# Choice of PDF set (NNPDF2.3 LO alphaS(mZ) = 0.13)
PDF:pSet = 13
```

Note that the NNPDF2.3 LO sets are provided for two values of the strong coupling,  $\alpha_s(M_Z) = 0.119$  and  $0.130$ ; we use the latter here. The sets have also been extended in order to have a wider validity range, in particular they are



**Fig. 13** Comparison of the gluon PDF at  $Q^2 = 2 \text{ GeV}^2$  between recent LO and LO\* PDF determinations. For NNPDF2.3LO, results for both  $\alpha_s(M_Z) = 0.130$  and  $\alpha_s(M_Z) = 0.119$  are shown

valid down to  $x = 10^{-9}$  and  $Q = 1 \text{ GeV}^2$ , precisely with the motivation of using them in LO event generators.

In Fig. 13, we compare the gluon PDF  $xg(x, Q^2)$  for the two NNPDF2.3 LO fits (central values only) with other recent LO and LO\* PDFs. There is a significant spread between the various LO/LO\* PDF determinations, reflecting the substantial theoretical uncertainties in LO fits. These differences are further enhanced at small  $x$  due to the lack of experimental constraints in this region. For instance, the CTEQ LO sets have a smaller gluon at small  $x$  than the other sets. The NNPDF2.3 LO PDF set for  $\alpha_s(M_Z) = 0.130$  is the largest at small  $x$ , beginning in  $x \sim 5 \times 10^{-6}$ , and is smaller than the other sets in the middle- $x$  region. These differences will translate into different phase-space populations for the multiparton-interaction processes relevant for the tuning of event generators.

### 3.2 The strong coupling and total cross sections

For hard QCD matrix elements in PYTHIA (including those for MPI), we use the same strong-coupling value as in the PDF set,<sup>11</sup>  $\alpha_s(M_Z) = 0.130$ :

```
SigmaProcess:alphaSvalue = 0.130
MultipartonInteractions:alphaSvalue = 0.130
```

This is slightly lower than the current default value of  $\alpha_s(M_Z) = 0.135$ , which however tends to produce too high inclusive jet rates, cf. the MCPLOTS web site [25]. Reducing the  $\alpha_s$  value also for MPI seems a reasonable first assumption;

<sup>11</sup> The difference between this  $\alpha_s$  value and that used for ISR/FSR will be discussed in Sect. 3.3.

it should result in a slightly less “jetty” underlying event, with activity shifted to lower  $p_{\perp}$  scales.

Already at this level, before considering any details of the MPI modelling, we can show one of the main theoretical reference distributions for multi-parton interactions: the integrated partonic QCD  $2 \rightarrow 2$  cross section (integrated above some  $p_{T\text{min}}$  scale), as a function of  $p_{T\text{min}}$ . All that is required to compute this are the PDFs, the value of  $\alpha_s(M_Z)$ , and the simple QCD LO  $d\sigma_{2 \rightarrow 2}$  differential cross sections. There is no dependence on other model parameters at this stage. Due to the  $1/p_T^4$  singularity of the differential Rutherford cross section,<sup>12</sup> this distribution diverges at low  $p_{T\text{min}}$ , an effect which is further amplified by the running of  $\alpha_s$  (which blows up at low scales) and the PDFs (which become large at low  $x$ ). MPI models reconcile the calculated divergent parton-parton cross section with the measured (or parametrized) total inelastic hadron-hadron cross section, by interpreting the divergence as a consequence of each hadron-hadron collision containing several parton-parton ones, with

$$\langle n \rangle_{\text{MPI}}(p_T \geq p_{T\text{min}}) \approx \frac{\sigma_{2 \rightarrow 2}(p_T \geq p_{T\text{min}})}{\sigma_{\text{inel}}} \quad (7)$$

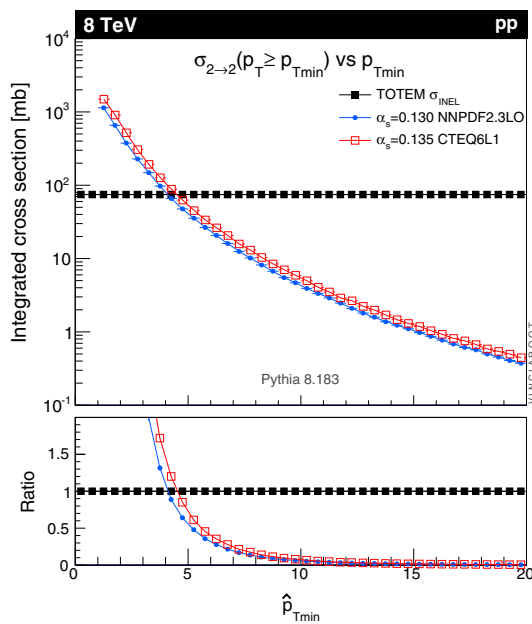
Note that there is some ambiguity whether to normalize to the total inelastic cross section, or to a diffraction-subtracted smaller number. To be conservative, we show a comparison to the full  $\sigma_{\text{inel}}$  in Fig. 14. We compare two different  $\alpha_s$  and PDF settings, corresponding to the choices made in the Monash 2013 tune (filled blue dots) and the current default 4C tune (open red squares), to the highly precise measurement of the total inelastic cross section at 8 TeV by the TOTEM collaboration [72],

$$\sigma_{\text{inel}}(8 \text{ TeV}) = (74.7 \pm 1.7) \text{ mb}. \quad (8)$$

For reference, the value obtained from the default Donnachie–Landshoff and Schuler–Sjöstrand parametrizations currently used in PYTHIA ( $\propto s^{0.0808}$  at high energies [73, 74]) is 73 mb, consistent with the TOTEM measurement.<sup>13</sup> The fact that the curves cross each other at a value of  $p_{T\text{min}} \sim 5 \text{ GeV}$  means that we can make a relatively model-independent statement that every inelastic event will, on average, contain at least one 5-GeV partonic subprocess. (This value agrees with that found by earlier analyses [76–78]). The

<sup>12</sup>  $t$ -channel gluon exchange gives an amplitude squared proportional to  $1/t^2$ , which for small  $p_T$  goes to  $1/p_T^4$ .

<sup>13</sup> We note, however, that the value obtained for the 8-TeV elastic cross section in PYTHIA is 20 mb, whereas the value measured by TOTEM is  $27.1 \pm 1.4 \text{ mb}$  [72]. While this discrepancy does not influence the normalization or modelling of inelastic events and hence is a non-issue in that context, an update of the total cross-section expressions in PYTHIA may be timely in the near future, e.g. using the updated Donnachie–Landshoff analysis in [75]. We also note that the decomposition of the inelastic cross section into individual non-diffractive and diffractive components, which follows Schuler–Sjöstrand [74], may also be due for an update.



**Fig. 14** Integrated LO QCD  $2 \rightarrow 2$  cross section vs  $p_{Tmin}$  for 8 TeV  $pp$  collisions, with two different  $\alpha_s$  and PDF choices, compared with the measured  $\sigma_{inel}$

corresponding  $p_{Tmin}$  scales at  $\sqrt{s} = 200$  or  $900$  GeV are just 1–2 GeV (see plots included Appendix B.3), hence the expected presence of “semi-hard” partonic substructure, at a scale of 5 GeV, in min-bias events is a qualitatively new feature at LHC energies; for completeness the corresponding scale at the Tevatron was about 2.5 GeV [77]. The plots in appendix B.3 also show extrapolations to higher energies. At 100 TeV, we expect the partonic cross section to saturate the total inelastic one at a  $p_T$  scale of 10 GeV.

### 3.3 Initial-state radiation and primordial kT

We follow the approach of the Perugia tunes of PYTHIA 6 [6, 8] and use the same  $\alpha_s(M_Z)$  value for initial-state radiation as that obtained for final-state radiation. That is, we use one-loop running with  $\alpha_s(M_Z) = 0.1365$  for both FSR and ISR. This choice is made essentially to facilitate matching applications, see e.g. [21]. Nonetheless, we emphasize that we do not regard this choice as mandatory, for the following reasons.

Firstly, since each collinear direction is associated with its own singular (set of) diagram(s), one can consistently associate at least the collinear radiation components with separate well-defined  $\alpha_s$  values without violating gauge invariance. Secondly, while the LO splitting functions for ISR and FSR are identical, they differ at higher orders (beyond the shower accuracy), and there are important differences between the collinear (DGLAP) evolution performed in PDF fits and the (coherent, momentum-conserving) evolution performed by

parton showers; these differences could well be desired to be reflected in slightly different effective scale choices for ISR with respect to FSR, one possibility then being to absorb this in a redefinition of the effective value of  $\alpha_s(M_Z)$ . Thirdly, and perhaps most importantly, while we agree that maintaining separate  $\alpha_s$  values (equivalent to making slightly different effective scale choices) for ISR and FSR is ambiguous for wide-angle radiation, we emphasize that merely using the same  $\alpha_s(M_Z)$  value for the two algorithms does *not* remove this fundamental ambiguity. This is because, in the context of a shower algorithm, the value of the renormalization scale depends upon *which* parton is branching, and that assignment is fundamentally ambiguous outside the collinear limit. For instance, an emitted gluon with a certain momentum will have a different  $p_\perp$  with respect to the beam (ISR), than it will with respect to a final-state parton (FSR), and hence the argument of  $\alpha_s$ , typically taken to be proportional to some measure of  $p_\perp$ , will be different, depending on who the emitter was. This effect is present in all parton-based shower algorithms and is *not* cured by arbitrarily setting  $\alpha_s(M_Z)$  to be the same for ISR and FSR. Using the same  $\alpha_s(M_Z)$  for both ISR and FSR (as we do here) should therefore not be perceived of as being more rigorous than not doing so; it is a choice we make purely for convenience. (The situation is slightly better in antenna-based showers [79–81], where there is no distinction between radiator and recoiler in the soft limit, hence the renormalization-scale choice is unique, at leading colour.)

The difference between the value  $\alpha_s(M_Z) = 0.130$  used for QCD matrix elements (and in the PDF evolution) and that used for ISR/FSR may be interpreted as follows. The former is specified in the  $\overline{MS}$  scheme, while the effective ISR/FSR one should presumably be interpreted in something closer to the so-called MC (CMW) scheme [28]. Taking the translation into account (corresponding roughly to a factor 1.6 on the value of  $\Lambda_{QCD}$ ), the PDF value comes out slightly lower than the shower one. Given the ambiguities caused by the non-identical nature of PDF and shower evolutions, however, we nonetheless regard this small difference as acceptable, in particular since the shower evolution is intrinsically somewhat slower than the PDF one, due to coherence effects and a more restrictive phase space that are not taken into account in the PDF evolution. For completeness, we note that the renormalization scale for ISR in PYTHIA is [27]:

$$\text{ISR: } \mu_R^2 = p_{\perp \text{Levol}}^2 = (1 - z)Q^2, \tag{9}$$

with  $Q^2 = -p^2$  the virtuality of the (spacelike) emitting parton (defined so that  $Q^2$  is positive; note that  $Q^2 = -p^2 + m_0^2$  is used for  $g \rightarrow Q\bar{Q}$  splittings) and  $z$  the energy fraction appearing in the DGLAP splitting kernels,  $P(z)$ , which in PYTHIA is defined as the ratio of  $\hat{s}$  values before and after the branching in question. (To estimate the shower uncertainties associated with this choice of renormalization scale, we recommend using  $\ln(\mu_R^2) \pm \ln(2)$ , corresponding to a factor

$\sqrt{2}$  variation of  $\mu_R$ , similarly to what was recommended for final-state radiation in Sect. 2.)

The remaining settings for the ISR evolution are taken over from the previous default tune. The relevant parameters in the code are:

```
# ISR: Strong Coupling (same as FSR)
SpaceShower:alphaSValue = 0.1365
SpaceShower:alphaSuseCMW = off
SpaceShower:alphaSOrder = 1
# ISR: Infrared Cutoff (fixed value at 2.0 GeV)
SpaceShower:samePTasMPI = off
SpaceShower:pT0Ref = 2.0
SpaceShower:ecmRef = 7000.0
SpaceShower:ecmPow = 0.0
# ISR: Coherence and Spin Correlations
SpaceShower:rapidityOrder = on
SpaceShower:phiPolAsym = on
SpaceShower:phiIntAsym = on
```

We choose a fixed ISR cutoff, rather than one that scales with CM energy, in order to maintain a correspondence between the ISR cutoff and the “primordial  $k_T$ ” component which parametrizes additional non-perturbative and/or unresolved motion in the beam remnant. This latter component does not scale with the CM energy (though it may depend on the  $Q^2$  scale of the hard process), hence we believe it is most consistent to keep the ISR cutoff fixed as well. Since we choose an ISR cutoff of 2 GeV (see the ISR parameter list above), there are no perturbative (ISR) corrections generated below that scale, and soft processes involving momentum transfers less than 2 GeV do not receive any perturbative corrections at all. To represent the combined effects of unresolved radiation and non-perturbative Fermi motion, we add a Gaussian-distributed primordial- $k_T$  component to the partons extracted from the proton at the low- $Q$  end of the ISR cascade. In the Monash tune, the width of the Gaussian starts at 0.9 GeV, for an infinitely soft process, and gradually rises to an asymptotic value of 1.8 GeV, with a characteristic “half-scale” of  $Q = 1.5$  GeV:

```
BeamRemnants:primordialkTsoft = 0.9
BeamRemnants:primordialkThard = 1.8
BeamRemnants:halfScaleForKT = 1.5
```

The half-scale of  $Q = 1.5$  GeV was chosen in order to prevent the primordial- $k_T$  component from generating momentum kicks larger than that of the “hard” process, for low-scale processes. The asymptotic value of 1.8 GeV was chosen by comparing to the  $p_\perp$  spectrum of the lepton pair in  $pp \rightarrow Z \rightarrow \ell^+\ell^-$  events measured by the ATLAS and CDF experiments [82, 84]. Note that PYTHIA’s parton shower is automatically corrected to reproduce the full LO  $Z + \text{jet}$  matrix element [27, 85], in a manner highly similar to (but predating) that of POWHEG [86]. Our value for primordial  $k_T$  (1.8 GeV) is slightly lower than the current default (2 GeV) and gives a better agreement with the low- $p_\perp$  part

of the lepton-pair  $p_\perp$  spectrum, as is illustrated in Fig. 15, for 7 TeV (top row) and 1800 GeV (bottom row)  $pp$  ( $p\bar{p}$ ) collisions. Note that the left-hand panes show a “closeup” of the peak region at low  $p_\perp$  while the right-hand panes show the full spectrum. (Note also that these  $p_\perp$  spectra are normalized to unity, so the normalization of the inclusive  $Z$  cross section drops out.)

In the ATLAS spectra, the feature around  $p_\perp^{\mu\mu} \sim 35$  GeV is repeated by all MCs in the comparisons shown on the MCPLOTS web site [25], hence we regard it as an artifact of the data. We note however that there is a tendency for PYTHIA to overshoot the data between  $p_\perp$  values of roughly 20 to 100 GeV, at both CM energies. This is an interesting region intermediate between low- $p_\perp$  bremsstrahlung and high- $p_\perp$   $Z$ +jet processes, which will be particularly relevant to reconsider in the context of matrix-element corrections at the  $\mathcal{O}(\alpha_s^2)$  level and beyond [87].

### 3.4 Minimum bias and underlying event

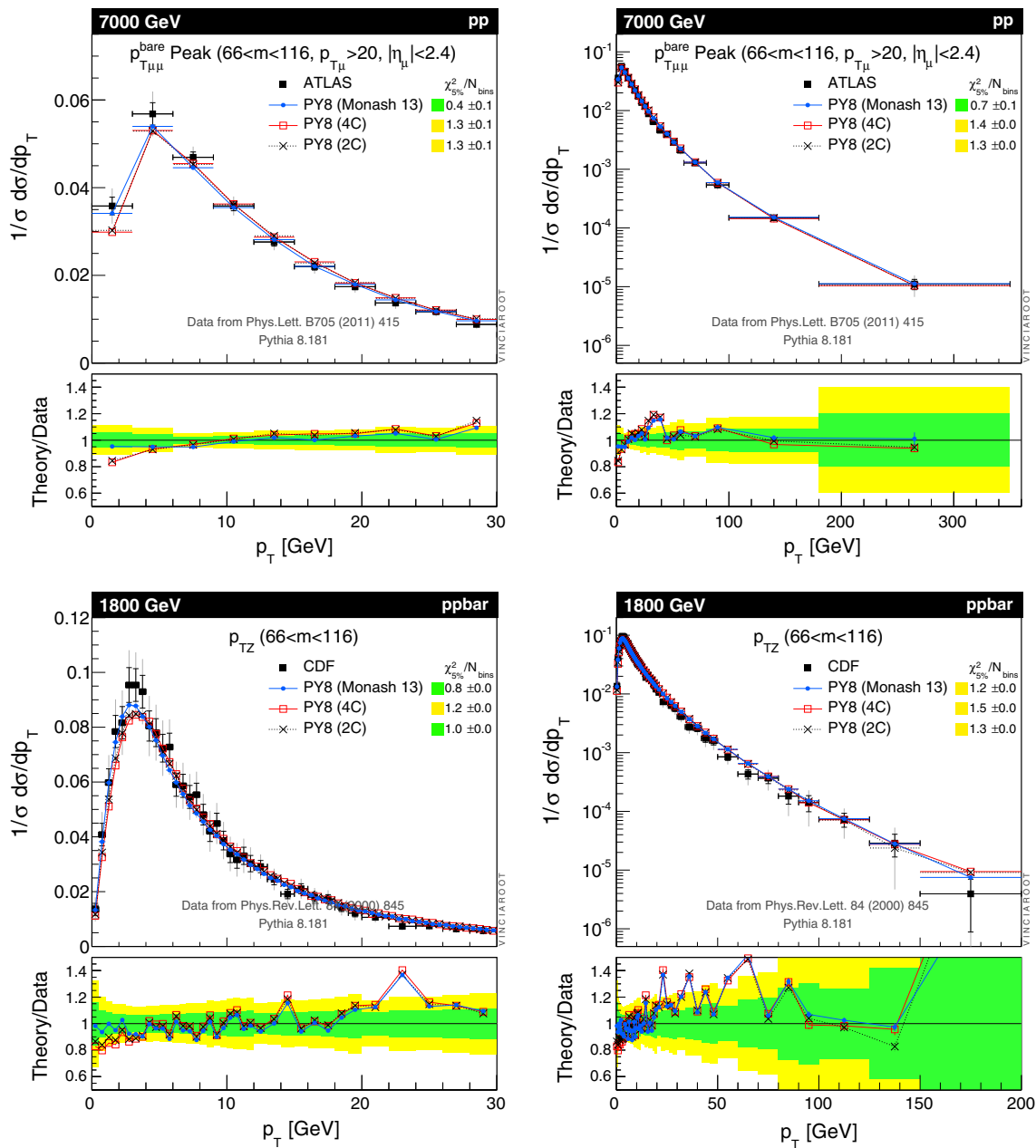
The Monash 2013 tune has been constructed to give a reasonable description of both soft-inclusive (“minimum-bias”) physics as well as underlying-event (UE) type observables. The difference between the two is sensitive to the shape of the hadron-hadron overlap profile in impact-parameter space (the UE probes the most “central” collisions while min-bias (MB) is more inclusive) and to the modeling of colour reconnections (CR). Most previous tunes, including the current default Tune 4C [9], have used a Gaussian assumption [76] for the transverse matter distribution, but this appears to give a slightly too low UE level (for a given average MB level).

For the Monash tune, we have chosen a slightly more peaked transverse matter profile [27], thus generating a relatively larger UE for the same average MB quantities. We note, however, that there are still several indications that the dynamics are not well understood, in particular when it comes to very low multiplicities (overlapping with diffraction), very high multiplicities (e.g., the so-called CMS “ridge” effect [88]), and to identified-particle spectra (e.g., possible modifications by re-scattering [89], string boosts from colour reconnections [90], or other collective effects).

For the 7-TeV reference energy we focus on here (energy scaling will be studied in the following subsection), the relevant parameters in the code are:

```
# Hadron transverse mass overlap density
profile
MultipartonInteractions:bProfile = 3
MultipartonInteractions:expPow = 1.85
# IR regularization scale for MPI and energy
scaling
MultipartonInteractions:pT0Ref = 2.28
MultipartonInteractions:ecmRef = 7000.
MultipartonInteractions:ecmPow = 0.215
```



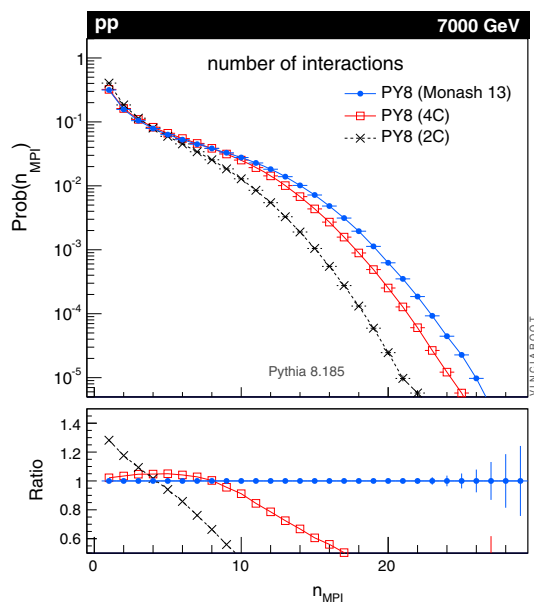


**Fig. 15** The peak (left) and tail (right) of the  $Z p_{\perp}$  distribution, as measured at 7 TeV (using “bare” muon pairs) [82] and 1.8 TeV (corrected to unphysical generator-level, see [83]) [84]

The slightly more peaked matter distribution, combined with a relatively low  $p_{\perp 0}$  value, produces an intrinsically broader distribution in the number of parton-parton interactions (MPI), illustrated by the theory-level plot in Fig. 16.

The sampling of the PDFs by MPI initiators (including also the hardest scattering in our definition of “MPI”), as a function of parton  $x$  values, is illustrated in Fig. 17, for the three tunes considered in this paper. The top left-hand pane shows the most inclusive quantity, simply the probability distribution of the  $x$  value of all MPI initiators (again, we

emphasize that we include the hardest-interaction initiators in our definition of “MPI” here), on a logarithmic  $x$  axis. Here we see that the NNPDF tune has a harder distribution both at large and small  $x$  as compared to the CTEQ6L1 tunes. The effect is particularly marked at small  $x$ . Since MPI is dominated by the low- $Q$  gluon PDF, cf. Fig. 12, this is precisely what we expect; the shape of the distribution of sampled  $x$  values follows that of the PDFs themselves. Indeed, the NNPDF2.3 gluon is harder than the CTEQ6L one for  $x > 0.2$  and for  $x < 10^{-5}$ .



**Fig. 16** *pp* collisions at 7 TeV. Number of MPI in inelastic events.

The relative dominance of the gluon PDF is illustrated by the bottom right-hand pane of Fig. 17, showing the gluon fraction (relative to all MPI initiators) as a function of  $\log_{10}(x)$ . Below  $x \sim 0.1$ , the NNPDF sampling is 80 % gluon-dominated, and the gluon fraction is higher than in CTEQ6L1 for both very small  $x < 10^{-5}$  as well as for very large  $x > 0.2$ .

A further consistency check is provided by the  $\bar{u}/u$  ratio, shown in the bottom left-hand pane of Fig. 17. This is consistent with unity (as expected for sea quarks) in the entire small- $x$  region  $x < 10^{-2}$ . The valence bump appears to be slightly more pronounced in the NNPDF tune (relative to the sea), since the  $\bar{u}/u$  ratio drops off more quickly above  $10^{-2}$ . This trend persists until the very highest bin, at  $x \sim 1$ , where the experimental uncertainties are extremely large. The CTEQ6L1 parametrization there forces the  $\bar{u}$  PDF to zero, while the NNPDF parametrization allows for a small amount of  $\bar{u}$  to remain even at the largest  $x$  values, though we note that they are still outnumbered by  $u$  quarks at a level of hundred-to-one.

The last pane of Fig. 17 shows the amount of  $x$  remaining in the beam remnant, after all MPI (including both the hardest interaction and additional MPI) have been considered, i.e.,

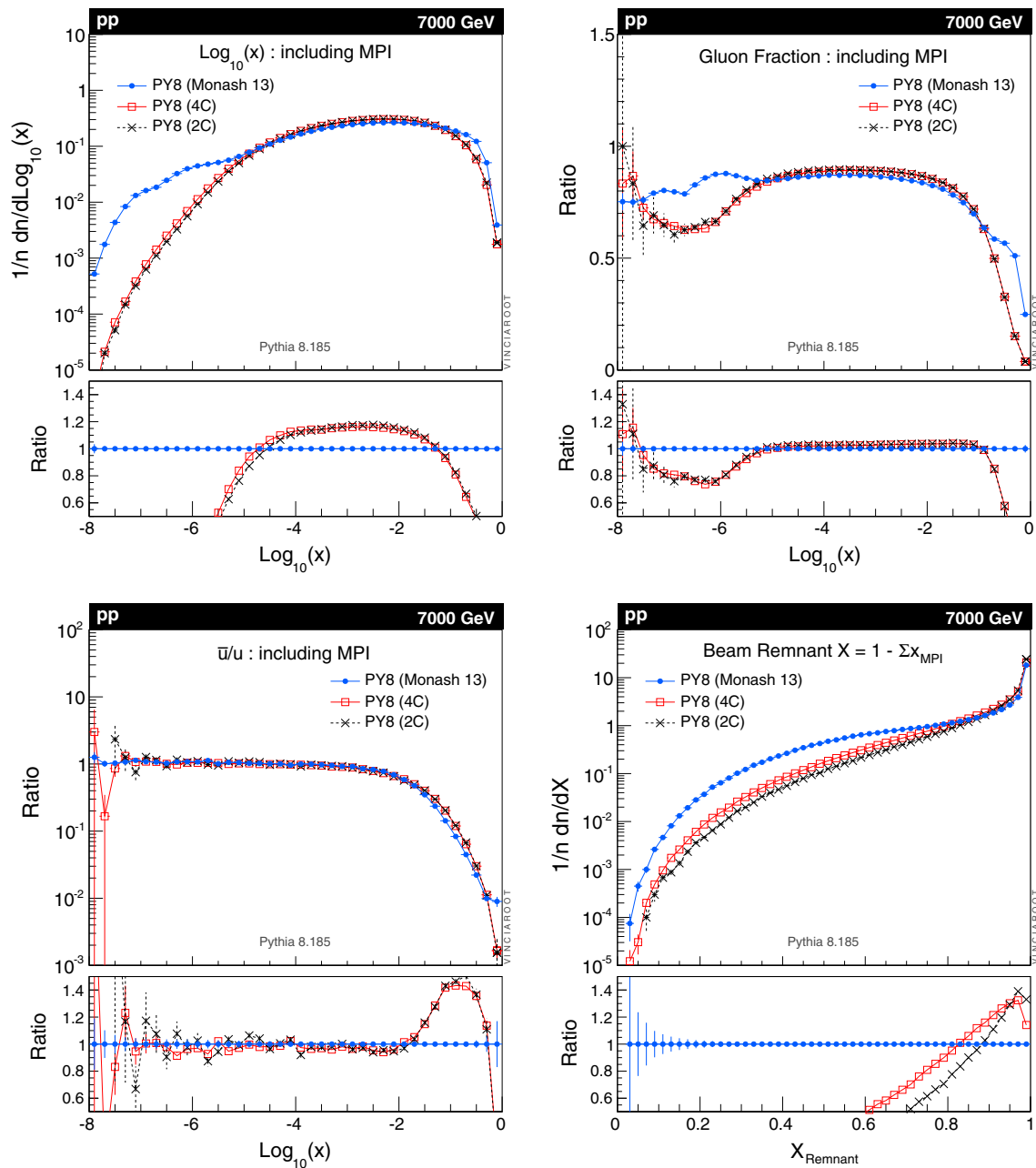
$$X_{\text{rem}} = 1 - \sum_{i \in \text{MPI}} x_i. \quad (10)$$

Note the linear scale in  $x$  on this plot, and the highly logarithmic axis. In the vast majority of cases, the beam remnant thus still retains over 90 % of the initial hadron energy. But there is a class of events, at the level of  $10^{-4}$  or  $10^{-5}$  of the total cross section (depending on the tune), in which the beam remnant retains less than 10 % of the incoming hadron energy.

Experiments studying the amount and distribution of forward scattered energy in particular may be able to tell us about whether this class of events, which we term “Catastrophic Energy Loss” events, really exists, and at what level. Note that these events are typically *not* caused by a single hard partonic scattering process, due to the high penalty associated with accessing PDFs in the region  $x > 0.5$ . Rather, they are an intrinsic consequence of MPI. A straightforward extrapolation, requiring a catastrophic energy loss on both sides of the event—more than 90 % of the energy scattered out of *both* beams, which we term “Total Inelastic Scattering”—may occur at a level of  $10^{-10}$ – $10^{-8}$  of the cross section, or between 10–1000 pb (though we of course only have PYTHIA’s word for it). This would be an extremely interesting part of hadron-hadron collision physics to study, very far from the single-interaction dominated limit, and hence potentially very sensitive to the existence of possible collective effects. Designing efficient triggers for this class of events would be a great accomplishment.

Turning now to physics distributions in min-bias events, the broader MPI distribution in the Monash tune translates to a broader charged-multiplicity spectrum, though the effect is modulated by the colour-reconnection model. The resulting multiplicity and  $p_{\perp}$  spectra are shown in Fig. 18, for “standard” fiducial cuts (top row:  $p_{\perp} \geq 500$  MeV,  $|\eta| < 2.5$ ,  $n_{\text{Ch}} \geq 1$ ) and “soft” fiducial cuts (bottom row:  $p_{\perp} \geq 100$  MeV,  $|\eta| < 2.5$ ,  $n_{\text{Ch}} \geq 2$ ), with the latter representing the most inclusive phase-space region accessible with the ATLAS detector. For both of the  $n_{\text{Ch}}$  distributions, we note that a significant “double-crested wave” pattern is still present in the ratio panes, though it has been dampened slightly. The  $p_{\perp}$  spectra in the right-hand panes are a bit below the data for the standard fiducial cuts and above it for the soft cuts, hence we regard the Monash tune as a reasonable compromise.

Pseudorapidity distributions are shown in Fig. 19. However, due to the complicated interplay between diffractive contributions at low multiplicity and high-multiplicity multiparton interactions (with associated questions of transverse matter density profile and colour reconnections), the average multiplicity by itself is a very difficult quantity to extract reliable conclusions from. Note also that the CMS measurement [92] shown in the top pane of Fig. 19 was corrected to an unphysical “non-single diffractive” event definition which essentially amounts to switching off single-diffractive contributions in the MC generator. (We note that later CMS measurements instead use a physical observable related to the diffractive mass to define NSD.) For the comparisons to CMS NSD data shown here, the single-diffractive contributions were switched off in the generator. With these caveats in mind, we note that both the 4C and Monash 2013 tunes are in good agreement with the CMS measurement, with the Monash one giving a slightly lower central charged-track



**Fig. 17** PDF sampling by MPIs in inelastic non-diffractive  $pp$  collisions at 7 TeV. *Top left* The  $x$  distribution of all MPI initiators (including the hardest scattering). *Top right* The fraction of MPI initiators which

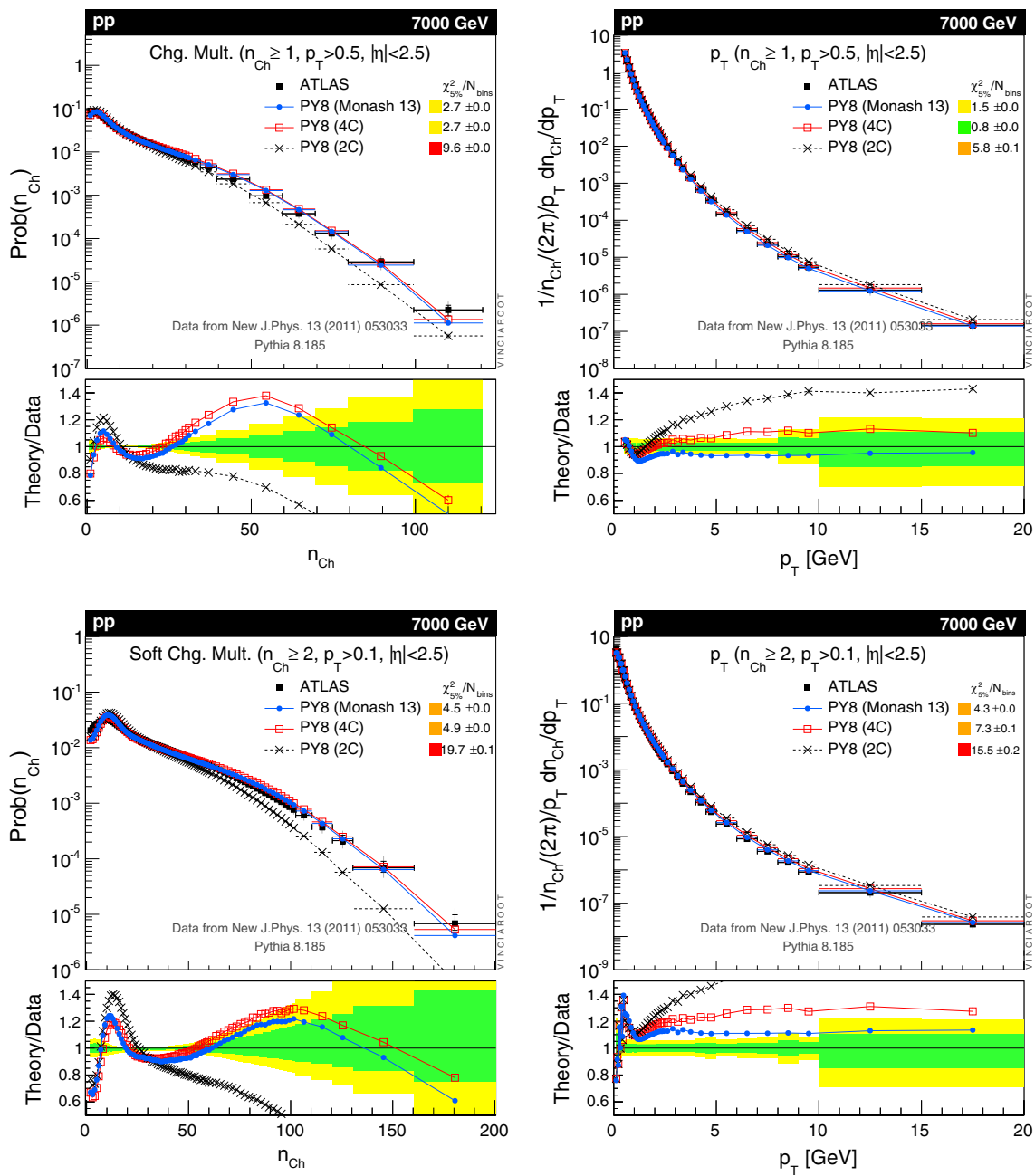
are gluons, as a function of  $x$ . *Bottom left* The  $\bar{u}/u$  ratio. *Bottom right* The distribution of the amount of  $x$  left in the beam remnant, after MPI (note linear scale in  $x$ )

density (by about 5 %). This is closer to the values observed in data, though as already noted in Sect. 1.1 we do not regard differences at the 5 % level as significant.

In the bottom two panes of Fig. 19, we focus on the forward region (with physical event selections). In particular, we see that the NNPDF set [20] generates a broader rapidity spectrum, so that while the activity in the central region (top pane) is reduced slightly, the activity in the very forward region actually increases, and comes into agreement with the TOTEM measurement [94], covering the range

$5.3 < |\eta| < 6.4$ . The bottom right-hand pane shows the forward energy flow measured by CMS [93], in the intermediate region  $3.23 < |\eta| < 4.65$ . The dependence on  $\eta$  is a bit steeper in the Monash tune than in the previous one, and more similar to that seen in the data.

A complementary observable, which is highly sensitive to interconnection effects between the MPI (and hence, e.g., to the effects of “colour reconnections” [95]), is the average charged-particle  $p_{\perp}$  as a function of the number of charged particles. In a strict leading-colour picture, each MPI would

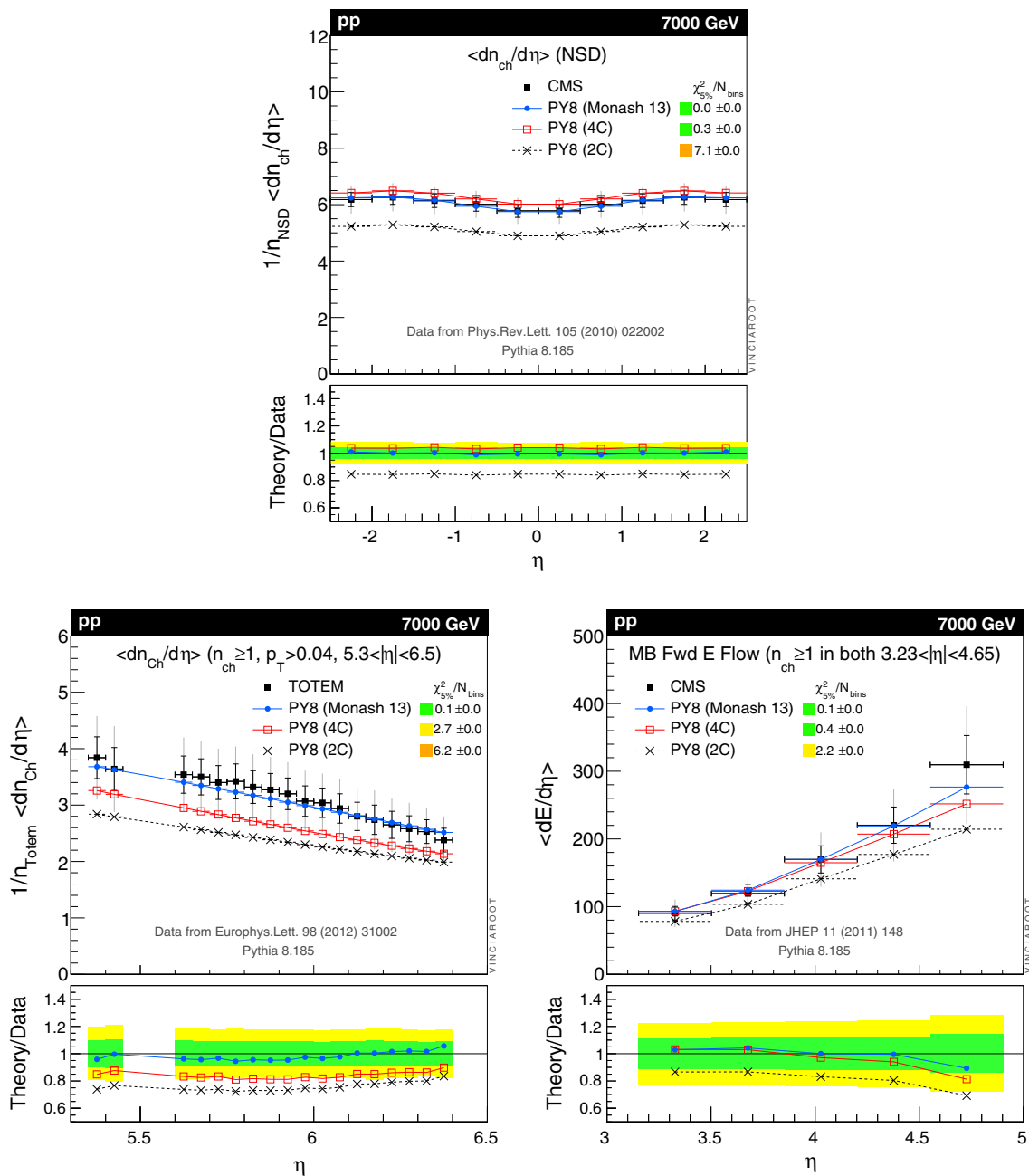


**Fig. 18** Min-bias  $pp$  collisions at 7 TeV. Charged-multiplicity and  $p_{\perp}$  distributions, with standard (*top row*) and soft (*bottom row*) fiducial cuts, compared to ATLAS data [91]

cause one or two new strings to be stretched between the remnants, but each such string would be independent (modulo endpoint effects); therefore (modulo jets) the  $p_{\perp}$  spectrum of the hadrons produced by each of these strings would be independent of the number of strings. The result would be a flat ( $p_{\perp}$ ) ( $n_{\text{Ch}}$ ) spectrum. Jets and colour reconnections both produce a rising spectrum. The spectra observed by ATLAS [91] are compared to the Monash, 2C, and 4C tunes in Fig. 20, for standard (left) and soft (right) fiducial cuts. Both of the Monash and 4C tunes reproduce the data quite

well, with  $\chi^2_{5\%}/N_{\text{bins}} < 1$ , while the older tune 2C had a higher CR strength optimized to describe Tevatron data [96]. We certainly consider the energy scaling of the effective CR strength among the most uncertain parameters of the current min-bias/underlying-event modelling (a similar conclusion was reached for the CR modelling in PYTHIA 6 in [97]), and intend to study the physics aspects of this issue more closely in a forthcoming paper.

For a more differential look at the event structure, we consider the charged-track  $\Delta\phi$  distributions with respect to the

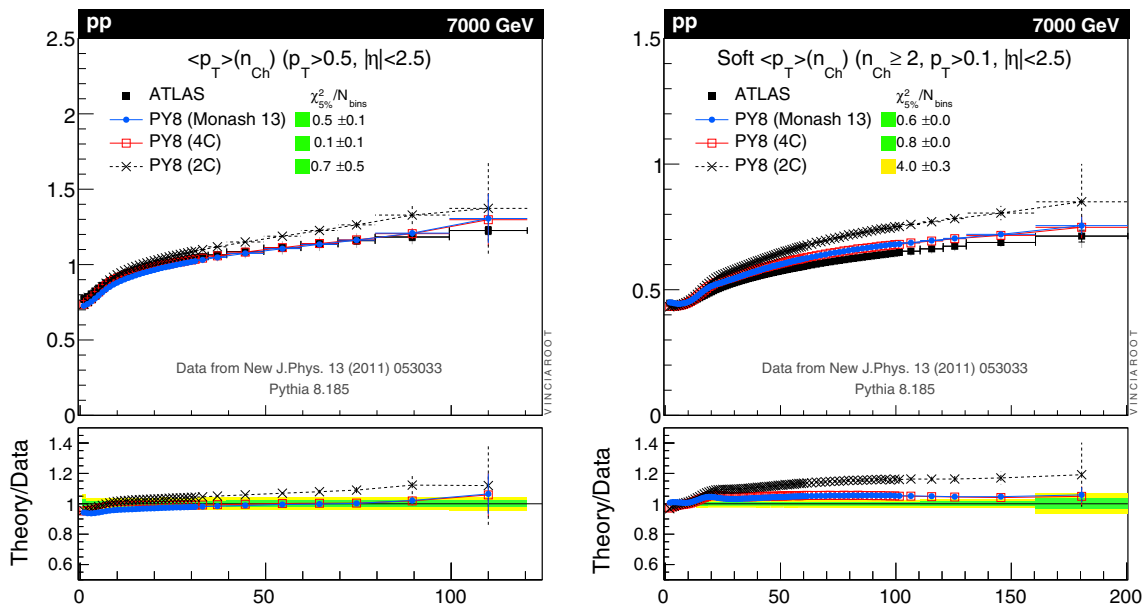


**Fig. 19** Charged-particle pseudorapidity distributions and forward energy flow in min-bias  $pp$  collisions at 7 TeV, compared to CMS [92,93] and TOTEM [94] data

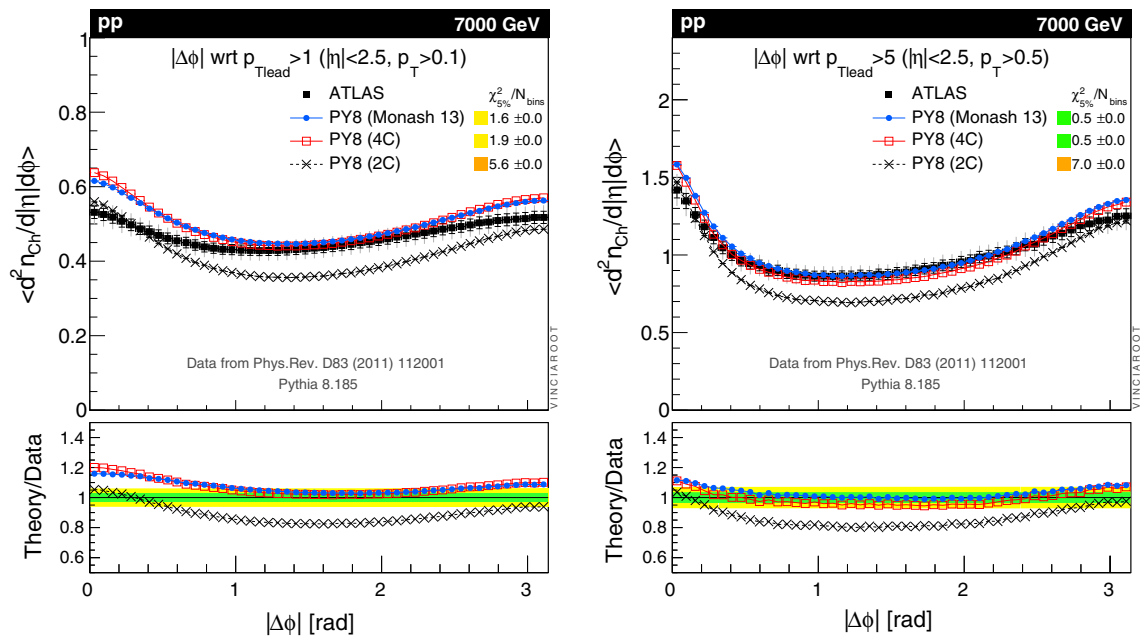
azimuthal angle of the leading track, in Fig. 21, compared with ATLAS data [98]. The plot in the left-hand pane corresponds to a requirement of  $p_{\perp lead} \geq 1$  GeV, while the one in the right-hand pane is for a harder trigger,  $p_{\perp lead} \geq 5$  GeV. The former can roughly be taken as characteristic of min-bias events, while the latter is related to the differential distribution of the underlying event. In both cases, the activity in the wide-angle region near  $\pi/2$  is significantly better described by the 4C and Monash 2013 tunes (which agrees with their improved description of the overall activity), while there is

a too strong peaking at low  $\Delta\phi$ , especially for the lowest  $p_{\perp lead}$  cut (left), possibly indicating that the structure of the min-bias events is still slightly too “lumpy” (i.e., jetty). For the higher  $p_{\perp lead}$  cut (right), the overcounting at very low  $\Delta\phi$  is already significantly milder, and we observe a good agreement with the data.

Turning now to the underlying event (UE), what matters most for high- $p_{\perp}$  jet studies is that the MC models describe the UE contamination per  $\Delta R$  jet area. The most important UE observable from this perspective is thus the  $p_{\perp}$  sum



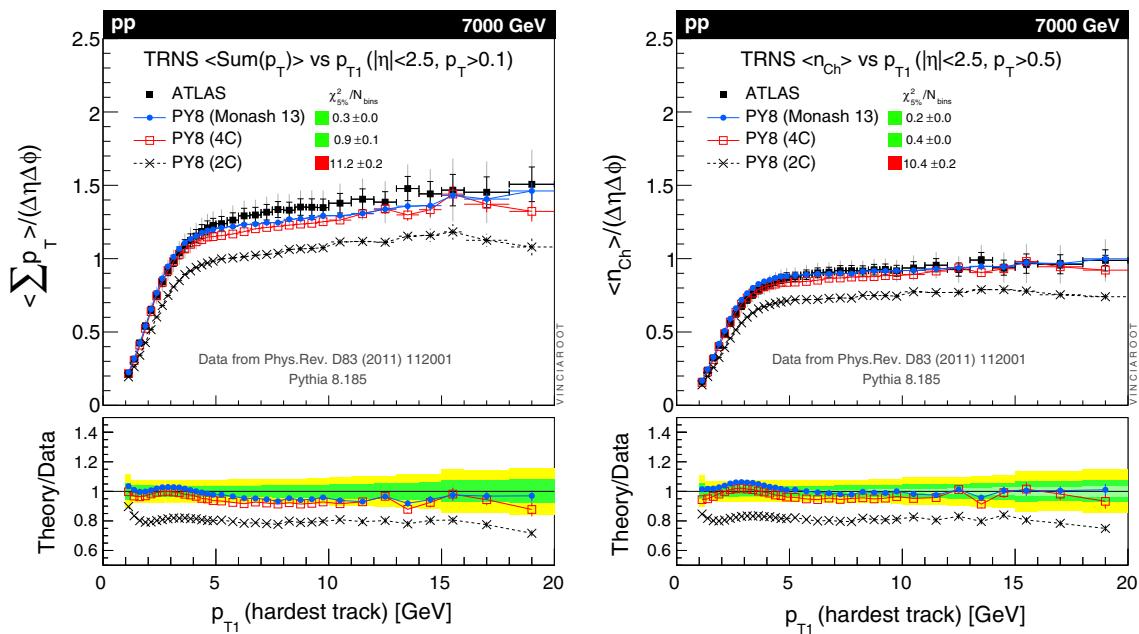
**Fig. 20** Average- $p_{\perp}$  vs. charged-multiplicity distributions in min-bias  $pp$  collisions at 7 TeV, with standard (left) and soft (right) fiducial cuts, compared to ATLAS data [91]



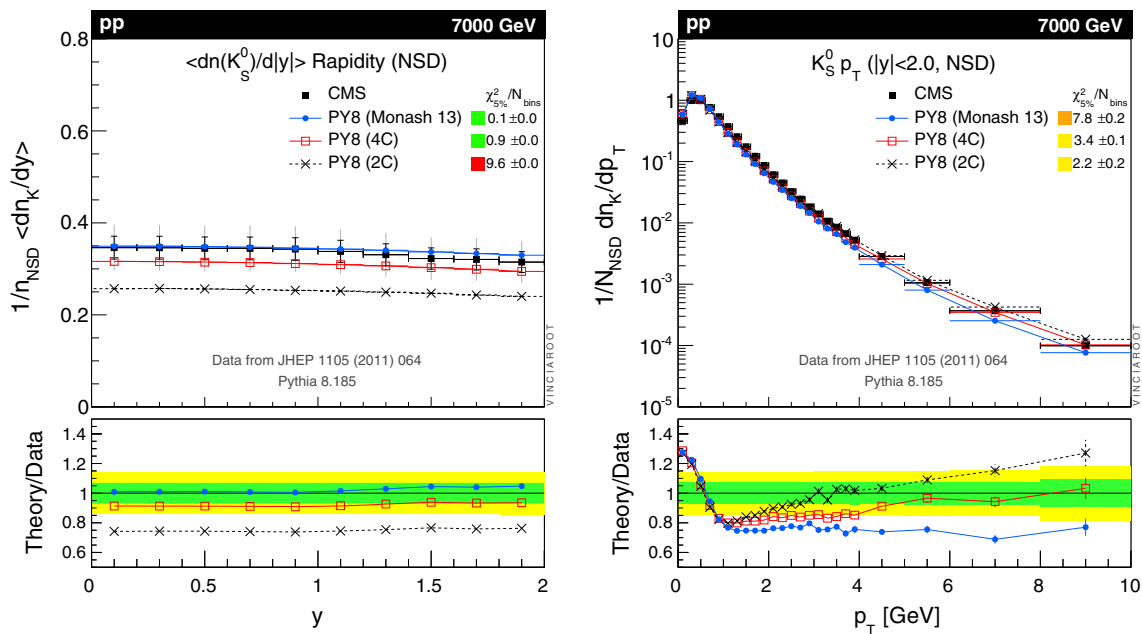
**Fig. 21**  $pp$  collisions at 7 TeV.  $\Delta\phi$  of charged particles with respect to the hardest track, for two different hardest-track triggers, compared with ATLAS data [98]

density in the UE, and its fluctuations. For charged particles at LHC, typically a  $p_{\perp}$  cut of 500 MeV is relevant, since softer tracks will form helices and hence not contribute to calorimetric jet energies. Neutral particles are of course relevant across all  $p_{\perp}$  scales. In Fig. 22, we show the charged  $p_{\perp}$  sum density (left, with the lowest possible  $p_{\perp}$  cut of 100 MeV) and the charged-track density (right, with a  $p_{\perp}$  cut of 500 MeV), in the so-called “Transverse Region”

(defined by  $60^{\circ} < \Delta\phi < 120^{\circ}$  with respect to the leading track), inside the ATLAS acceptance of  $|\eta| < 2.5$  [98]. As is now well known the Tevatron extrapolations (represented here by Tune 2C) predicted a UE level which was 10–20 % below the LHC data. Both the current default tune 4C (which included LHC data) and the Monash 2013 tune exhibit significantly better agreement with the LHC measurements, with the Monash one giving a slight additional improvement in the



**Fig. 22** *pp* collisions at 7 TeV. UE (“Transverse region”) transverse-momentum sum density (left) and charged-track density (right), compared with ATLAS data [98]



**Fig. 23** *pp* collisions at 7 TeV.  $K_S^0$  rapidity and  $p_{\perp}$  spectrum, compared with CMS data [99]

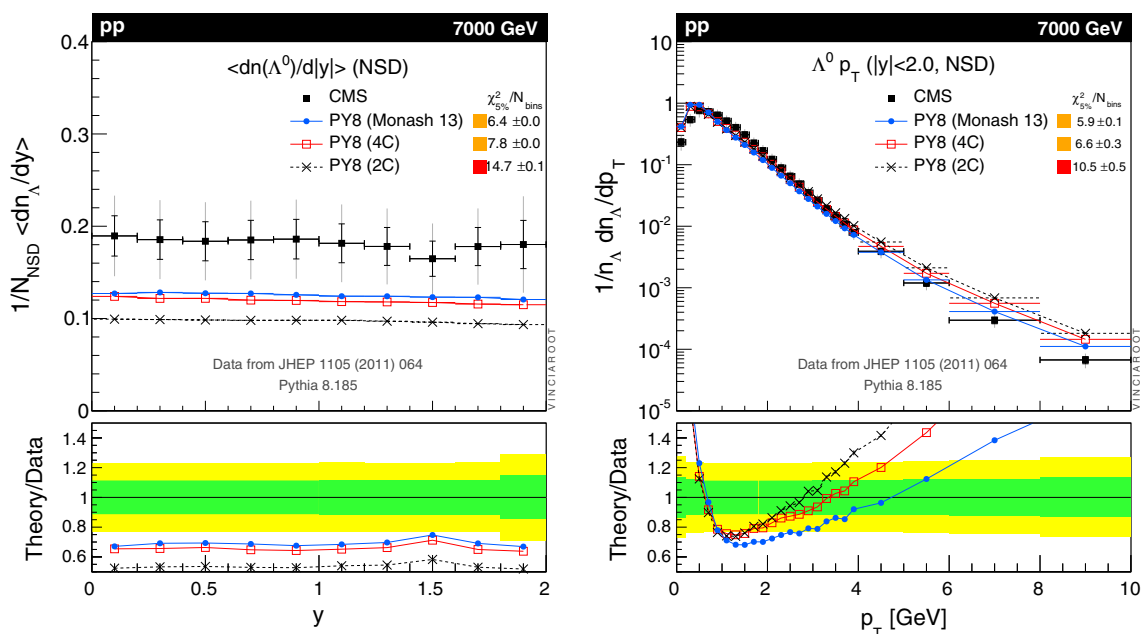
$\chi^2_{5\%}$  values. We conclude that the Monash 2013 tune parameters are appropriate for both min-bias and UE studies.

### 3.5 Identified particles at LHC

While the description of inclusive charged particles, discussed in the previous section, is acceptable, larger discrepan-

cies emerge when we consider the spectra of identified particles. We here focus on strange particles, in particular  $K_S^0$  mesons and  $\Lambda^0$  hyperons in Figs. 23 and 24, respectively. The experimental measurements come from CMS [99]. Additional comparisons to strange-particle spectra ( $K^*$ ,  $\phi$ , and  $\Xi$ ) are collected in Appendix B.2.

In the  $K_S^0$  rapidity distribution, shown in the left-hand pane of Fig. 23, we observe that tune 4C exhibits a mild under-



**Fig. 24**  $pp$  collisions at 7 TeV.  $\Lambda^0$  rapidity and  $p_{\perp}$  spectrum, compared with CMS data [99]

production, of about 10 %. Though it might be tempting to speculate whether this could indicate some small reduction of strangeness suppression in  $pp$  collisions, however, we already noted in Sect. 2.1 that the strangeness production in  $ee$  collisions also needed to be increased by about 10 %. After this adjustment, we see that the overall  $K_S^0$  yield in the Monash 2013 tune is fully consistent with the CMS measurement. Nonetheless, we note that the momentum distribution is still not satisfactorily described, as shown in the right-hand pane of Fig. 23. Our current best guess is therefore that the overall rate of strange quarks is consistent, at least in the *average* min-bias collision (dedicated comparisons in high-multiplicity samples would still be interesting), but that the phase-space distribution of strange hadrons needs more work. Similarly to the case in  $ee$  collisions, cf. Fig. 6, the model predicts too many very soft kaons, though we do not currently know whether there is a dynamic link between the  $ee$  and  $pp$  observations.

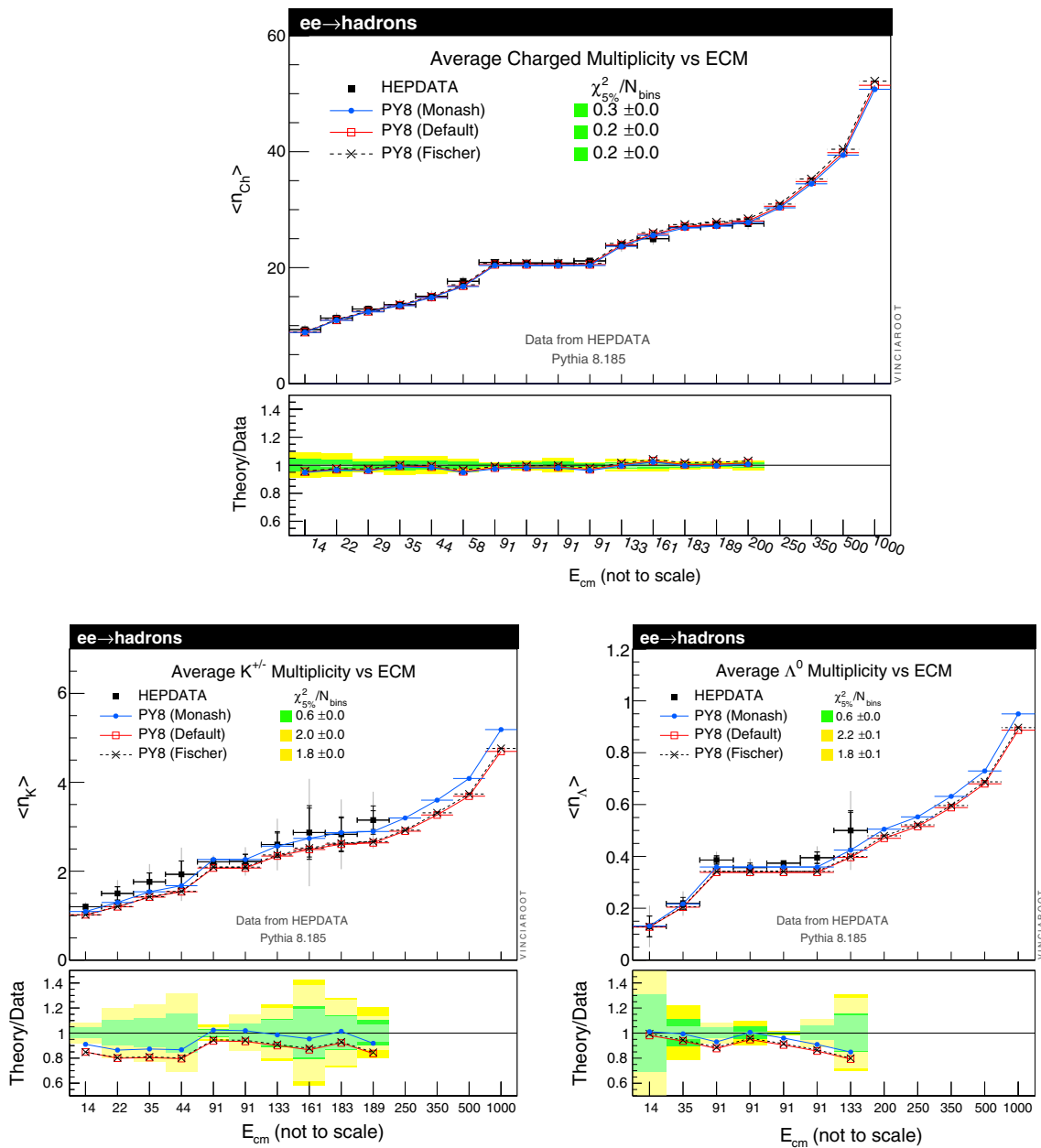
For strange baryons, we note that the increase in the  $\Lambda^0$  fraction in  $ee$  collisions (cf. Fig. 5) does *not* result in an equivalent improvement of the  $\Lambda^0$  rate in  $pp$  collisions, shown in Fig. 24. The Monash 2013 tune still produces only about 2/3 of the observed  $\Lambda^0$  rate (and just over half of the observed  $\Xi^-$  rate, cf. Appendix B.2). We therefore believe it to be likely that an additional source of net baryon production is needed (at least within the limited context of the current PYTHIA modelling), in order to describe the LHC data. The momentum spectrum is likewise quite discrepant, exhibiting an excess at very low momenta (stronger than that for kaons), a dip between 1–4 GeV, and then an excess of very hard  $\Lambda^0$  production. The latter hard tail is somewhat milder

in the Monash 2013 tune than previously, and it may be consistent with the trend also seen in the  $\Lambda^0$  spectrum at LEP, cf. Fig. 6. We conclude that baryon production still requires further modelling and tuning efforts.

#### 4 Energy scaling

Though energy scaling these days mostly refers to the scaling of  $pp$  collisions (see e.g., [97]), an important first step is to consider the scaling of observables in  $ee \rightarrow \gamma^*/Z \rightarrow$  hadrons. This scaling contains information on the relative contributions of perturbative and non-perturbative fragmentation. Thus, at low  $ee$  energies, the non-perturbative components of the fragmentation model dominate, while perturbative bremsstrahlung increases in importance towards higher  $ee$  energies. In Fig. 25, we consider the scaling of the average charged-particle multiplicity and that of charged Kaons and Lambda baryons from CM energies of 14 to 200 GeV, obtained from measurements available at HEPDATA. Below the Z pole, the measurements we include mostly come from TASSO [100], though a few points on  $\langle n_{Ch} \rangle$  come from HRS (at 29 GeV [101]) and TOPAZ (at 57.8 GeV [102]). At the Z pole, the data come from the four LEP experiments [38, 51–53], with the latter extending also to energies above  $M_Z$  [103–108]. For completeness and as reference for future  $ee$  collider studies, model extrapolations for CM energies up to 1000 GeV are also shown (though still only including the  $ee \rightarrow \gamma^*/Z \rightarrow$  hadrons component, as usual with photon ISR switched off).





**Fig. 25**  $e^+e^- \rightarrow$  hadrons. Energy scaling of  $\langle n_{Ch} \rangle$ ,  $\langle n_{K^\pm} \rangle$ , and  $\langle n_{\Lambda} \rangle$ , in  $e^+e^- \rightarrow q\bar{q}$  events, including comparisons to measurements from HEPDATA for CM energies from 14 to 200 GeV. Also shown are model extrapolations up to 1000 GeV

From the plots in Fig. 25, it is clear that there are no significant differences between the energy scaling of the three  $ee$  tunes considered here (mainly reflecting that they have been tuned to same reference point, at 91.2 GeV, and that their scaling is dictated by the same underlying physics model), and that their energy dependence closely matches that observed in data. However, the increased amount of non-perturbative strangeness production in the Monash tune leads to a better agreement with the overall normalization of the  $K^\pm$  and  $\Lambda$  rates at all energies.

Moving to  $pp$  collisions, the plots in Fig. 26 show the scaling of the average charged multiplicity (left column) and multiplicity distributions (right column) in min-bias collisions from 7000 GeV (top row) to 900 GeV (middle row) and 200 GeV (bottom row), compared with data from CMS [92,109], ATLAS [91], and UA5 [110,111]. We regret the omission of additional relevant min-bias measurements from the Tevatron and RHIC experiments here, but have chosen to focus in this paper mainly on the LHC. The comparisons at 7 TeV were already discussed in the main section on  $pp$  colli-

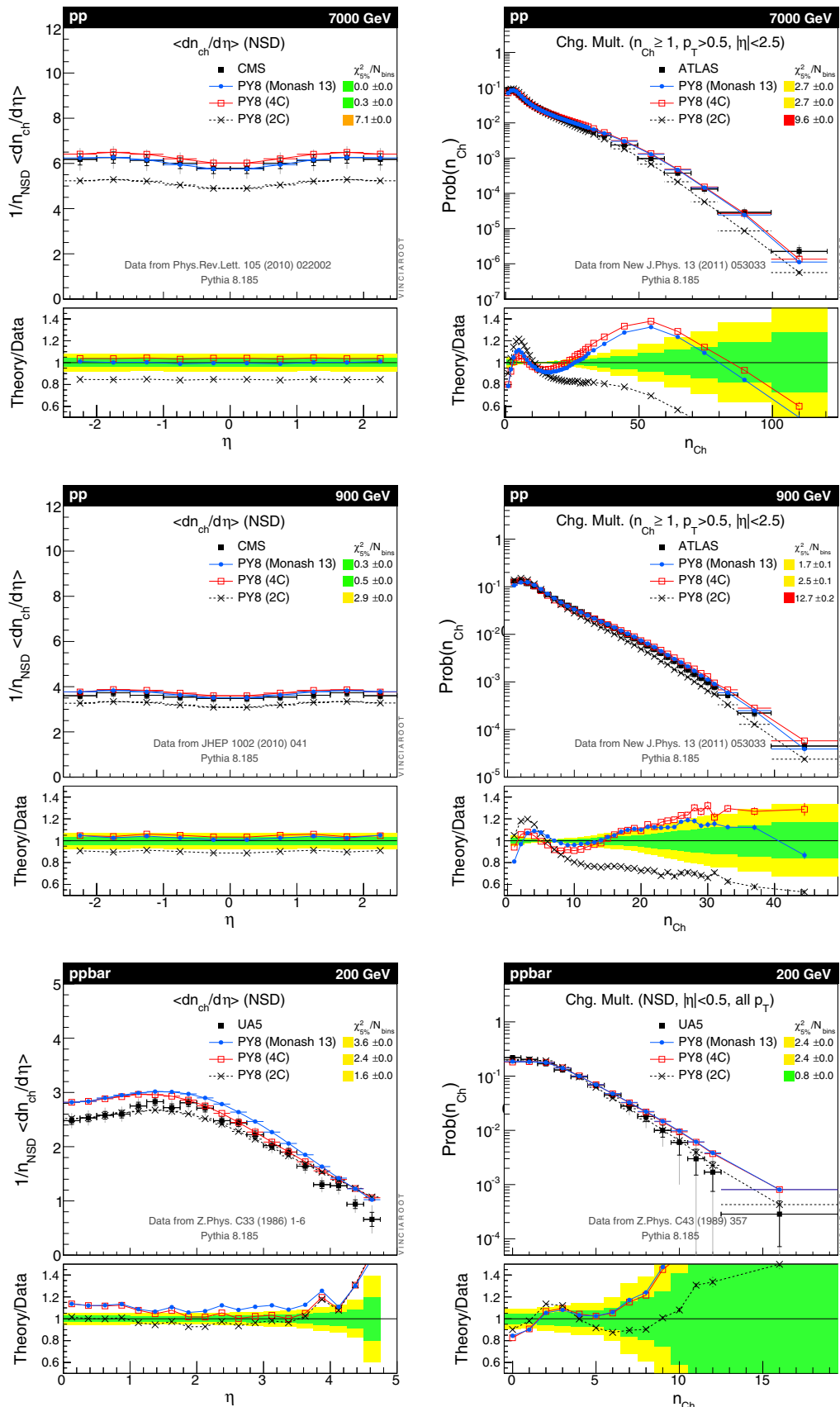


Fig. 26 Min-bias  $pp$  events, from 200 to 7000 GeV. Energy scaling of  $\langle dn_{Ch}/d\eta \rangle$  (left) and  $P(n_{Ch})$  (right)

sions, Sect. 3. At 900 GeV, the Monash 2013 tune again gives a roughly 5 % lower average central charged multiplicity than the 4C one, with a better description of the tail towards high multiplicities. At 200 GeV, the UA5 measurement we include here extends over the full rapidity and  $p_{\perp}$  range, hence the interplay between diffraction and low-multiplicity non-diffractive processes is presumably (much) more important. We believe imperfections in this modelling to be the likely cause of the significant discrepancies observed at high  $\eta$  and for  $n_{\text{Ch}} \leq 20$  at these energies. Since a dedicated study of this interplay is beyond the scope of this study, we limit ourselves merely to stating this observation, as a point for future studies to help clarify.

Finally, in Fig. 27, we compare to the underlying event measured in the highly useful energy scan that was performed at the Tevatron in the last days before its shutdown [22,23], during which extremely high min-bias statistics were collected at 300 and 900 GeV CM energy over a period of a few days. As was already noted in Sect. 3, the UE at 7 TeV is slightly larger in the Monash 2013 tune than in tune 4C. As can be seen from the plots here, the two tunes give comparable results for all the Tevatron energies. Interestingly, the UE plateau region at 900 and 1960 GeV is reached sooner in these models than in the data, translating to a roughly 10–20 % too low UE level for leading-track  $p_{\perp}$  values in the neighbourhood of the transition from the rise to the plateau (roughly for leading-track  $p_{\perp}$  values  $2 < p_{T1} < 10$  GeV). This indicates that the energy scaling of the UE modeling and in particular the details of its transition between central and peripheral collisions, is still not satisfactorily understood.

## 5 Conclusions and exhortation

We have presented a reanalysis of the constraints on fragmentation in  $ee$  collisions, and applied the results to update the final-state fragmentation parameters in PYTHIA 8. We combine these parameters with a tune to hadron-collider data, using a new NNPDF 2.3 LO QCD+QED PDF set, which has been encoded so it is available as an internal PDF set in PYTHIA 8, independently of LHAPDF [112].

In this PDF set as well as in our tune, the value of the strong coupling for hard-scattering matrix elements is fixed to be  $\alpha_s(M_Z) = 0.13$ , consistent with other LO determinations of it. For initial- and final-state radiation, our tune uses the effective value  $\alpha_s(M_Z) = 0.1365$ . The difference is consistent with an effective translation between the  $\overline{\text{MS}}$  and CMW schemes. We note that alternative (LO, NLO, and NNLO) NNPDF 2.3 QCD+QED sets with  $\alpha_s(M_Z) = 0.119$  are also available in the code, for people who want to check the impact of using a different  $\alpha_s(M_Z)$  value and/or higher-order PDF sets on hard-scattering events. For the purpose of such studies, we point out that it is possible, in PYTHIA 8, to preserve

most of the features of the shower- and underlying-event tuning by changing only the PDF for the hard-scattering matrix elements, leaving the PDF choice for the shower evolution and MPI framework unaltered (see the PYTHIA 8 HTML manual's PDF section, under `PDF:useHard`).

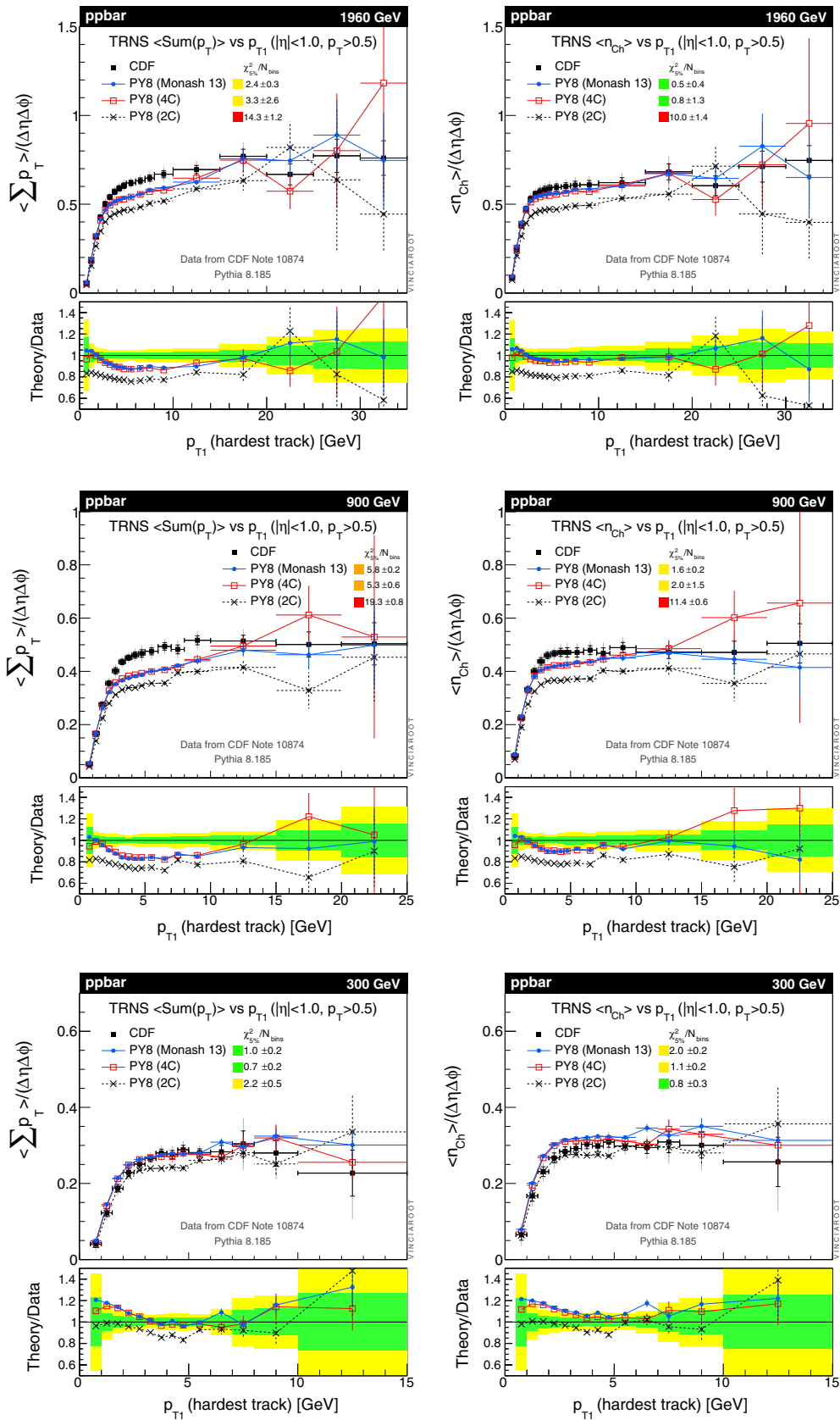
The updated parameters are available as an option starting from PYTHIA 8.185, by setting

```
Tune:ee = 7 and Tune:pp = 14.
```

By no means do we claim that this should be regarded as the final word in tuning the PYTHIA 8 Monte Carlo model. First of all, the model continues to evolve. For instance, developments foreseen for the near future include updates of colour reconstructions, diffraction, and the treatment of  $g \rightarrow q\bar{q}$  splittings. Any of these should in principle be accompanied by a reevaluation of the model constraints.

Moreover, despite the comprehensive view of collider data we have attempted to take in this study, there still remains several issues that were not addressed, including: initial-final interference and coherence effects [113,114] (probably more a modelling issue than a tuning one); reliable estimates of theoretical uncertainties [8,17,24,29,97,115]; diffraction<sup>14</sup> [74,116,117] and other colour-singlet phenomena such as onium production; long-distance (e.g., forward-backward, forward-central, and “ridge”-type) correlations [88,118–123];  $B$ -hadron decays [124]; and tuning in the presence of matrix-element matching, at LO and NLO (see [21,29,115,125,126] for recent phenomenological studies). Especially in the latter context of matrix-element matching, we expect that in many cases PYTHIA 8 will be used together with codes such as ALPGEN [127], MADGRAPH [128], aMCatNLO [129], or POWHEG [130], either using the matching algorithms of those programs themselves, or via any of PYTHIA's several internal (LHEF-based [131]) implementations of matching schemes (POWHEG [86], CKKW-L [132–134], MLM [135,136], UMEPS [137], NL3 [138], UNLOPS [139]). The impact of such corrections on MC tuning depends on the details of the matching scheme (especially its treatment of unitarity), and there is in general a non-negligible possibility of “mis-tuning” when combining a stand-alone tune with ME corrections. A simple example illustrating this is the effective value of  $\alpha_s(M_Z)$ , which for a leading-order tune is typically of order 0.13, while a consistent NLO correction scheme should be compatible with values closer to 0.12 [29]. There is also the question of the

<sup>14</sup> In particular, the constraints on fragmentation mainly come from SLD and LEP, where the non-perturbative parameters are clearly defined at the shower cutoff scale,  $Q_{\text{had}}$ , whereas diffraction is dominated by soft physics, for which the definition of the effective hadronization scale is less clear. The amount of MPI in hard diffractive events also requires tuning.



**Fig. 27** The Tevatron energy scan. The underlying event (*left* average summed- $p_{\perp}$  density and *right* average track density, in the transverse region, as function of leading-track  $p_{\perp}$ ) for  $p\bar{p}$  collisions at 300 (*bottom row*), 900 GeV (*middle row*), and 1960 GeV (*top row*)

running order of  $\alpha_s$ . The propagation of such changes from the level of hard matrix elements through the shower and hadronization tuning process are still not fully explored, and hence we advise users to perform simple cross-checks, such as checking the distributions presented in this paper, before and after applying matrix-element corrections. Parameters that appear on both sides of the matching, such as  $\alpha_s$ , should also be checked for consistency [21].

We noted several issues concerning the  $ee$  data used to constrain the fragmentation modelling, that it would be good to resolve. In particular, we find some tensions between the identified-particle rates extracted from (1) HEDPATA, (2) Sec. 46 of the PDG, and (3) the  $Z$  boson summary table in the PDG, as discussed in more detail in Sect. 2, and concerning which we made some (subjective) decisions to arrive at a set of hopefully self-consistent constraints for this work. We also note that the overall precision of the fragmentation constraints could likely be significantly improved by an FCC- $ee$  type machine, such as Tera- $Z$ , a possibility we hope to see more fully explored in the context of future  $ee$  QCD phenomenology studies.

We conclude that the new parameter set does improve significantly on the previous default values in several respects, including better agreement with data on:

1. the net strangeness fraction (has been increased by 10 %, reflected not only in improved kaon and hyperon yields, but also in the  $D_s$  and  $B_s$  fractions),
2. the ultra-hard fragmentation tail (has been softened, especially for leading baryons and for  $D$  and  $B$  hadrons),
3. the  $p_{TZ}$  spectrum (softened at low  $p_{TZ}$ ),
4. the minimum-bias charged multiplicity in the forward region (has increased by 10 %),
5. the underlying event at 7 TeV (is very slightly higher than before).

Some questions that remain open include the following. We see a roughly 20 % excess of very soft kaons in both  $ee$  and  $pp$  environments, cf. Figs. 6 and 23, despite the overall kaon yields being well described, and the overall baryon yields at LHC appear to be underestimated by at least 30 % despite good agreement at LEP. The momentum spectra of heavier strange particles are also poorly reproduced, in particular at LHC. It is interesting and exciting that some of the LHC spectra appear to be better described by allowing collective flow in a fraction of events (cf. the EPOS model [140]), though we believe the jury is still out on whether this accurately reflects the underlying physics. For instance, it has been argued that colour reconnections can mimic flow effects [90], and they may also be able to modify the yield of baryons if the creation/destruction of string junctions is allowed [141]. We look forward to future discussions on these issues.

We round off with an exhortation for follow-ups on this study to provide:

- **Not only central tunes:** experiments and other user-end colleagues need more than central descriptions of data; there is an increasing need for serious uncertainty estimates. In the context of tune variations, it is important to keep in mind that the modelling uncertainties are often intrinsically non-universal. Therefore, the constraints obtained by considering data uncertainties only (e.g., in the spirit of PROFESSOR's eigentunes [17]) can at most constitute a lower bound on the theoretical uncertainty (similarly to the case for PDFs). A serious uncertainty estimate includes some systematic modelling variation, irrespectively of, and in addition to, what data allows (e.g., in the spirit of the Perugia set of tunes for PYTHIA 6 [8]). We therefore hope the future will see more elaborate combinations of data- and theory-driven approaches to systematic tune uncertainties;
- **Not only global tunes:** the power of MC models lies in their ability to *simultaneously* describe a large variety of data, hence we do not mean to imply that one should give up on universality and tune to increasingly specific corners of phase space, disregarding (or de-emphasizing, with lower weights) all others. However, as proposed in [97], one can obtain useful explicit tests of the universality of the underlying physics model by performing independent tunes on separate “physics windows”, say in the forward vs. central regions, for different event-selection criteria, at different collider energies, or even for different collider types. In this connection, just making one global “best-fit” tune may obscure tensions between the descriptions of different complementary data sets. By performing independent tunes to each data set separately, and checking the degree of universality of the resulting parameters, one obtains a powerful cross check on the underlying physics model. If all sets produce the same or similar parameters, then universality is OK, hence a global tune makes very good sense, and the remaining uncertainties can presumably be reliably estimated from data alone. If, instead, some data sets result in significantly different tune parameters, one has a powerful indication that the universality of the underlying modelling is breaking down, which can lead to several productive actions: (1) it can be taken into account in the context of uncertainty variations, (2) the nature of the data sets for which non-universal tune parameters are obtained can implicitly indicate the nature of the problem, leading to more robust conclusions about the underlying model than merely whether a tune can/cannot fit the data, and (3) the observations can be communicated to the model authors in a more unambiguous way, hopefully resulting in a speedier cycle of model improvements.

We hope that the Monash 2013 tune parameters may serve as a useful starting point for phenomenology studies and for future PYTHIA 8 tuning efforts.

**Acknowledgments** We thank S. Mrenna, M. Ritzmann, and T. Sjöstrand for comments on the manuscript and L. de Nooij for pointing us to the ALICE  $K^*$  and  $\phi$  measurements [142] and to the ATLAS  $\phi$  measurements in [143]. The work of J. R. is supported by an STFC Rutherford Fellowship ST/K005227/1. This work was supported in part by the Research Executive Agency (REA) of the European Commission under the Grant Agreements PITN-GA-2012-315877 (MCnet).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

Funded by SCOAP<sup>3</sup> / License Version CC BY 4.0.

## Appendix A: Monash 2013 tune parameters

In Tables 3, 4, 5, 6 and 7, we list the FSR, fragmentation, parameters for the Monash tune of PYTHIA. For reference, we compare them to the current default parameters.

**Table 3** Final-state radiation (FSR) parameters

FSR parameters	Monash 13	(Default)	Comment
TimeShower:alphaSvalue	= 0.1365	= 0.1383	! Effective alphaS(mZ) value
TimeShower:alphaSorder	= 1	= 1	! Running order
TimeShower:alphaSuseCMW	= off	= off	! Translation from $\overline{MS}$ to CMW
TimeShower:pTmin	= 0.50	= 0.40	! Cutoff for QCD radiation
TimeShower:pTminChgQ	= 0.50	= 0.40	! Cutoff for QED radiation
TimeShower:phiPolAsym	= on	= on	! Asymmetric azimuth distributions

**Table 4** String-breaking parameters

HAD parameters	Monash 13	(Default)	Comment
# String breaks: pT and z distributions			
StringPT:sigma	= 0.335	= 0.304	! Soft pT in string breaks (in GeV)
StringPT:enhancedFraction	= 0.01	= 0.01	! Fraction of breakups with enhanced pT
StringPT:enhancedWidth	= 2.0	= 2.0	! Enhancement factor
StringZ:aLund	= 0.68	= 0.3	! Lund FF a (hard fragmentation supp)
StringZ:bLund	= 0.98	= 0.8	! Lund FF b (soft fragmentation supp)
StringZ:aExtraSquark	= 0.0	= 0.0	! Extra a when picking up an s quark
StringZ:aExtraDiquark	= 0.97	= 0.50	! Extra a when picking up a diquark
StringZ:rFactC	= 1.32	= 1.00	! Lund-Bowler c-quark parameter
StringZ:rFactB	= 0.855	= 0.67	! Lund-Bowler b-quark parameter
# Flavour composition: mesons			
StringFlav:ProbStoUD	= 0.217	= 0.19	! Strangeness-to-UD ratio
StringFlav:mesonUDvector	= 0.5	= 0.62	! Light-flavour vector suppression
StringFlav:mesonSvector	= 0.55	= 0.725	! Strange vector suppression
StringFlav:mesonCvector	= 0.88	= 1.06	! Charm vector suppression
StringFlav:mesonBvector	= 2.2	= 3.0	! Bottom vector suppression
StringFlav:etaSup	= 0.60	= 0.63	! Suppression of eta mesons
StringFlav:etaPrimeSup	= 0.12	= 0.12	! Suppression of eta' mesons
# Flavour composition: baryons			
StringFlav:probQQtoQ	= 0.081	= 0.09	! Diquark rate (for baryon production)
StringFlav:probSQtoQQ	= 0.915	= 1.000	! Strange-diquark suppression
StringFlav:probQQ1toQQ0	= 0.0275	= 0.027	! Vector diquark suppression
StringFlav:decupletSup	= 1.0	= 1.0	! Spin-3/2 baryon suppression
StringFlav:suppressLeadingB	= off	= off	! Optional leading-baryon suppression
StringFlav:popcornSpair	= 0.9	= 0.5	!
StringFlav:popcornSmeson	= 0.5	= 0.5	!

**Table 5** Parton-distribution (PDF) and matrix-element (ME) parameters

PDF and ME parameters	Monash 13	(Default)	Comment
PDF:pSet	= 13	= 8	! PDF set for the proton
SigmaProcess:alphaSvalue	= 0.130	0.135	! alphaS(MZ) for matrix elements
MultiPartonInteractions:alphaSvalue	= 0.130	0.135	! alphaS(MZ) for MPI

**Table 6** Initial-state radiation (ISR) and primordial- $k_T$  parameters

ISR parameters	Monash 13	(Default)	Comment
SpaceShower:alphaSvalue	= 0.1365	= 0.137	! Effective alphaS(mZ) value
SpaceShower:alphaSorder	= 1	= 1	! Running order
SpaceShower:alphaSuseCMW	= off	= off	! Translation from $\overline{MS}$ to CMW
SpaceShower:samePTasMPI	= off	= off	! ISR cutoff type
SpaceShower:pT0Ref	= 2.0	= 2.0	! ISR pT0 cutoff
SpaceShower:ecmRef	= 7000.0	= 1800.0	! ISR pT0 reference ECM scale
SpaceShower:ecmPow	= 0.0	= 0.0	! ISR pT0 scaling power
SpaceShower:rapidityOrder	= on	= on	! Approx coherence via y-ordering
SpaceShower:phiPolAsym	= on	= on	! Azimuth asymmetries from gluon pol
SpaceShower:phiIntAsym	= on	= on	! Azimuth asymmetries from interference
TimeShower:dampenBeamRecoil	= on	= on	! Recoil dampening in final-initial dipoles
BeamRemnants:primordialKTsoft	= 0.9	= 0.5	! Primordial kT for soft procs
BeamRemnants:primordialKTthard	= 1.8	= 2.0	! Primordial kT for hard procs
BeamRemnants:halfScaleForKT	= 1.5	= 1.0	! Primordial kT soft/hard boundary
BeamRemnants:halfMassForKT	= 1.0	= 1.0	! Primordial kT soft/hard mass boundary

**Table 7** Multi-parton-interaction (MPI), colour-reconnection (CR), and diffractive parameters

MPI parameters	Monash 13	(Default)	Comment
MultipartonInteractions:pT0Ref	= 2.28	= 2.085	! MPI pT0 IR regularization scale
MultipartonInteractions:ecmRef	= 7000.0	= 1800.0	! MPI pT0 reference ECM scale
MultipartonInteractions:ecmPow	= 0.215	= 0.19	! MPI pT0 scaling power
MultipartonInteractions:bProfile	= 3	= 3	! Transverse matter overlap profile
MultipartonInteractions:expPow	= 1.85	= 2.0	! Shape parameter
BeamRemnants:reconnectRange	= 1.8	= 1.5	! Colour Reconnections
SigmaTotal:zeroAXB	= on	= on	! Carried over from 4C
SigmaDiffractive:dampen	= on	= on	! Carried over from 4C
SigmaDiffractive:maxXB	= 65.0	= 65.0	! Carried over from 4C
SigmaDiffractive:maxAX	= 65.0	= 65.0	! Carried over from 4C
SigmaDiffractive:maxXX	= 65.0	= 65.0	! Carried over from 4C
Diffraction:largeMassSuppress	= 4.0	= 2.0	! High-mass diffraction suppression power

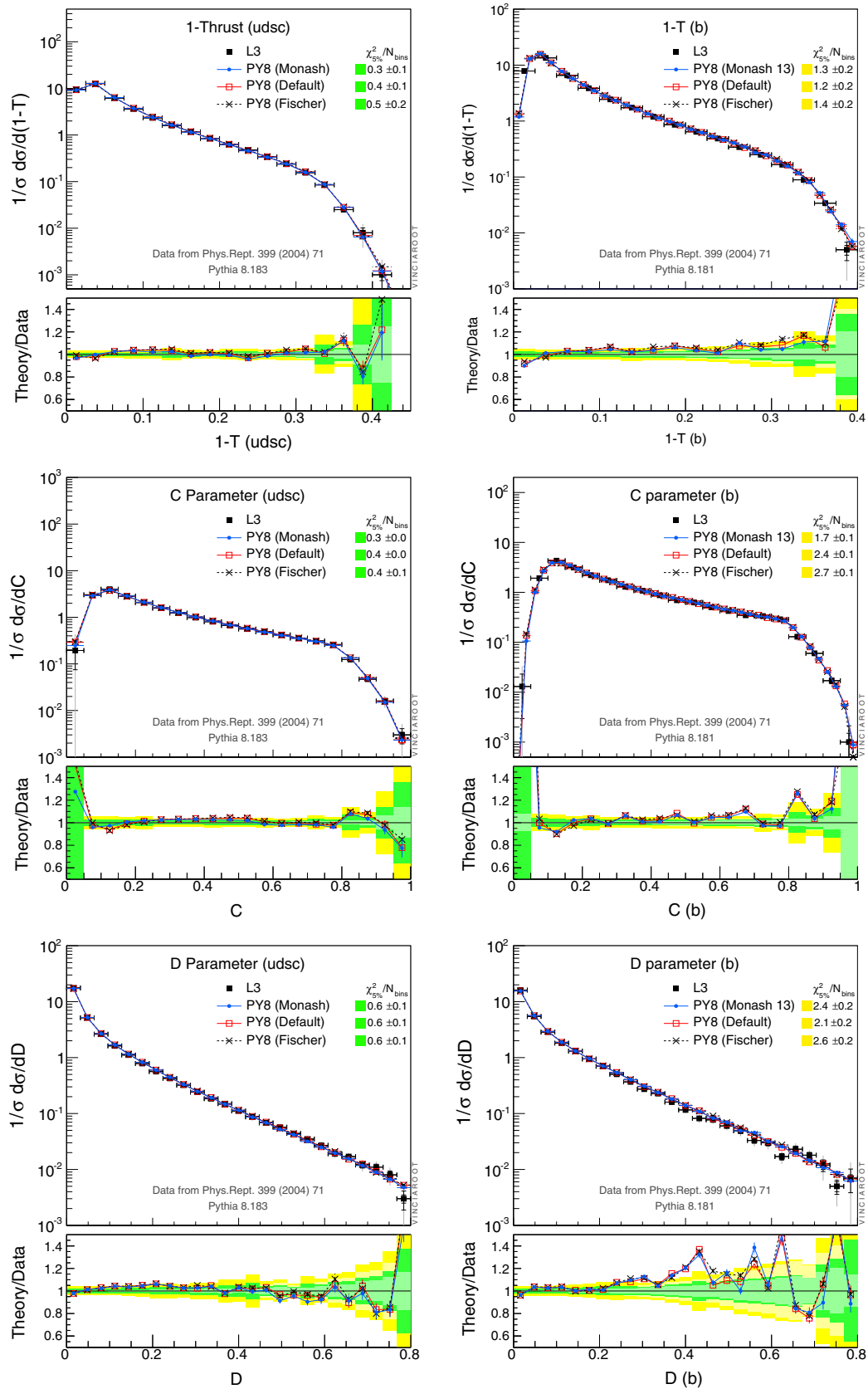
## Appendix B: Additional plots

### B.1 LEP event-shape distributions

To keep the main body of the paper as uncluttered as possible, we collect various plots of event-shape distributions

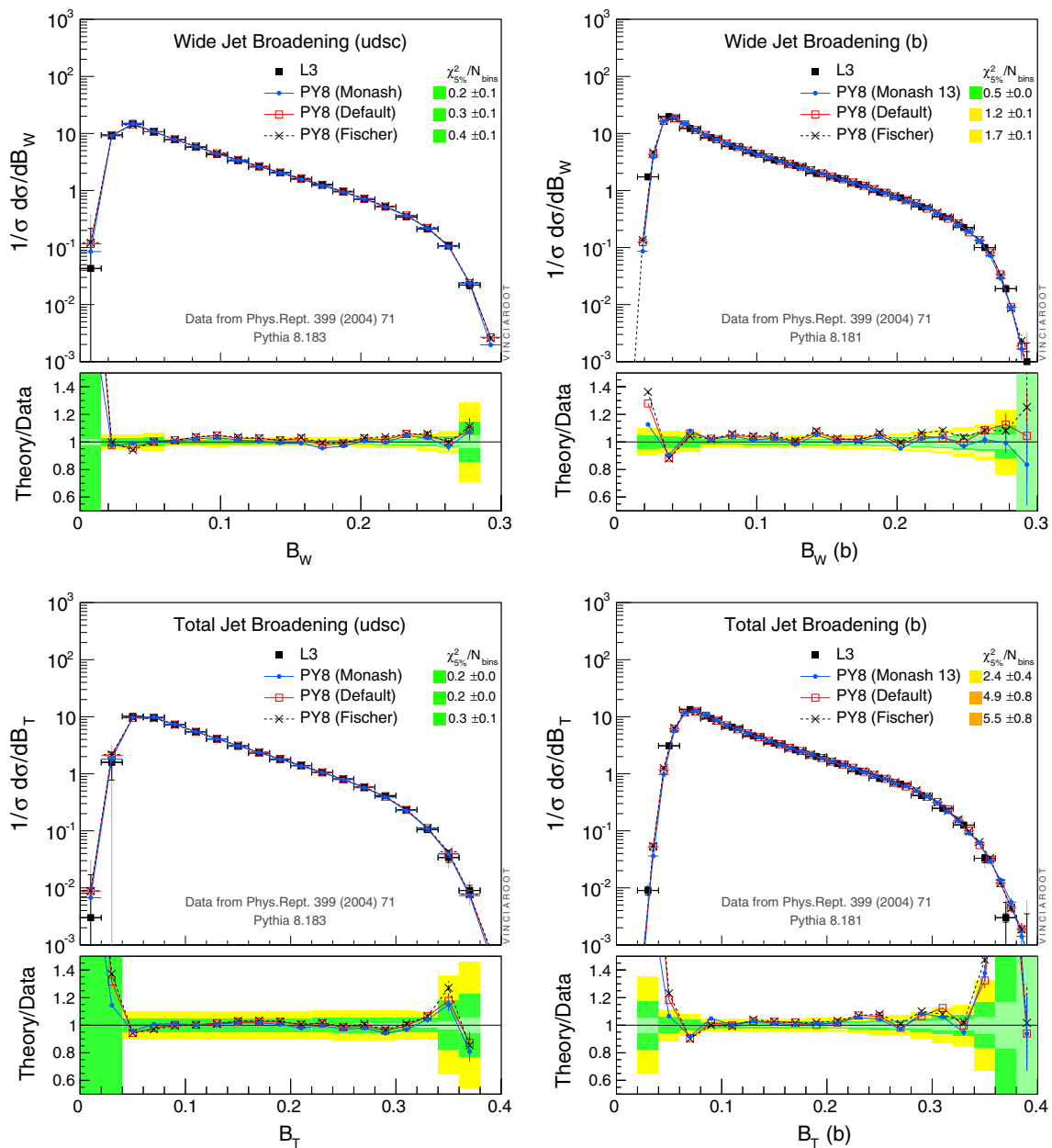
in Figs. 28 and 29, separated into light-flavour and  $b$ -tagged events on the left and right, respectively.

The experimental results come from the L3 experiment [26]. However, since the data points are only available with 3-digit precision, some of the least populated bins contain artifacts like uncertainties being reported as exactly zero, etc.



**Fig. 28** Hadronic *Z* decays at  $\sqrt{s} = 91.2$  GeV. The *T*, *C*, and *D* event-shape parameters, as measured by L3 [26], for light-flavour (*left*) and *b*-tagged (*right*) events, respectively





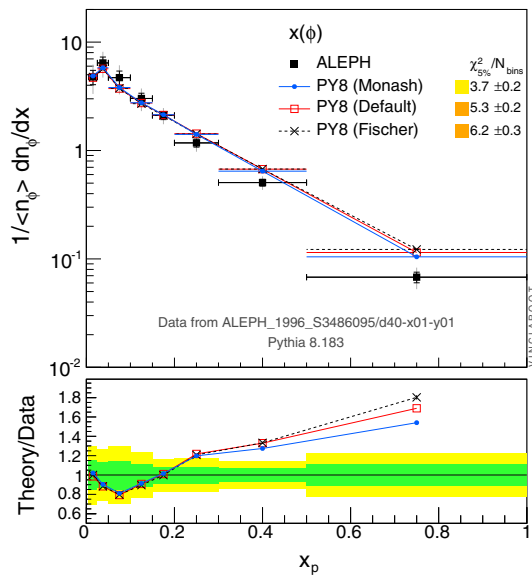
**Fig. 29** Hadronic Z decays at  $\sqrt{s} = 91.2$  GeV. The  $B_W$  and  $B_T$  event-shape parameters, as measured by L3 [26], for light-flavour (left) and b-tagged (right) events, respectively

Thus, we have been forced to make the following modifications to the data set.

The statistical uncertainty was reported as zero for the last two bins of light-flavour Thrust as well as for the last bin of the  $C$ ,  $D$ , and  $B_T$  parameters. Uncertainties  $< 10^{-3}$  were derived using an approximate statistical scaling based on the contents and uncertainties of the other bins. Likewise, the systematical uncertainty for the last bin of Thrust was given as zero, which we have replaced by the upper limit,

$5 \times 10^{-4}$ . The last bin of  $B_T$  quoted a measured  $y$  value of zero; removed in this study.

For the heavy-flavour tagged event shapes, more significant rounding issues were present. Thus, several of the first and last bins of each distribution either quoted zero (statistical and/or systematic) uncertainties, or ones with only a single digit of precision (such as 0.001, for which the rounding error could be up to  $\sim 50\%$ ). We have interpreted all such values conservatively, inserting by hand a fourth digit



**Fig. 30** Hadronic Z decays at  $\sqrt{s} = 91.2$  GeV.  $\phi$  meson  $x$  spectrum

on the uncertainties as large as could be consistent with rounding.

### B.2 Additional particle spectra

In addition to the  $K$  and  $\Lambda$  spectra shown in the main body of the paper (Sects. 2.2 and 3.5), we here include for reference the  $x$  spectra of  $\phi$  mesons, protons and  $\Xi$  baryons at LEP in Figs. 30 and 31, the  $p_T$  spectrum of  $K^*$  mesons and the

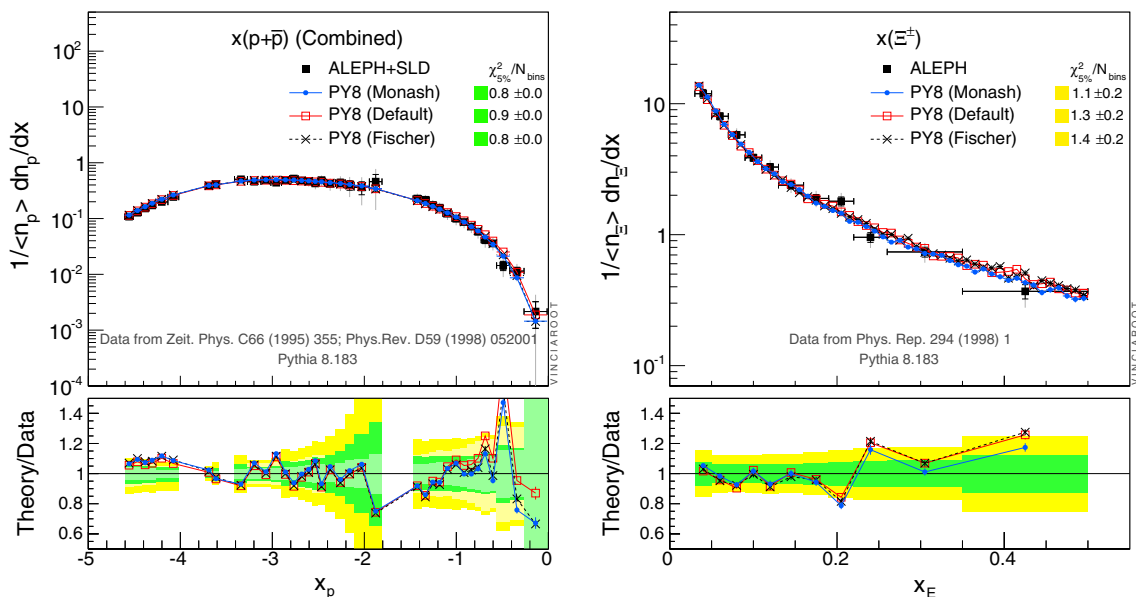
rapidity and  $p_T$  spectra of  $\phi$  mesons at LHC in Fig. 32 (with absolute normalizations, to the number of inelastic events), and the rapidity spectrum of  $\Xi$  baryons at LHC, in Fig. 33.

The transverse-momentum spectra of  $K^*$  and  $\phi$  mesons in Fig. 32 exhibit the same qualitative behaviour as that of the  $K_S$  mesons (Fig. 23), namely an excess at very soft momenta below  $\sim 500$  MeV and a depletion at slightly higher momenta between 1 and 2 GeV. As discussed in Sect. 3.5, we did not find a way to remove these undesirable features in the Monash 2013 tune, suggesting that this is an issue that further theoretical modeling will be needed to resolve.

The rapidity spectrum of  $\Xi$  baryons, Fig. 33, shows that, although the Monash tune does produce more  $\Xi$  baryons overall (as expected also from the relative increase of  $\Xi$  production at LEP, cf. Fig. 5), there is still a significant deficit of  $\Xi$  baryons at the LHC, almost a factor 2 compared with the data. This is qualitatively similar to the situation for  $\Lambda$  baryons (Fig. 24) discussed in Sect. 3.5. Since new physics mechanisms may be required to “explain” the missing baryons, we conclude that further measurements and better precision on both the  $\Lambda$  and  $\Xi$  sectors (in addition to any other baryons that may be accessible) would be highly interesting. More explicit recommendations can be found in Sects. 3.5 and 5.

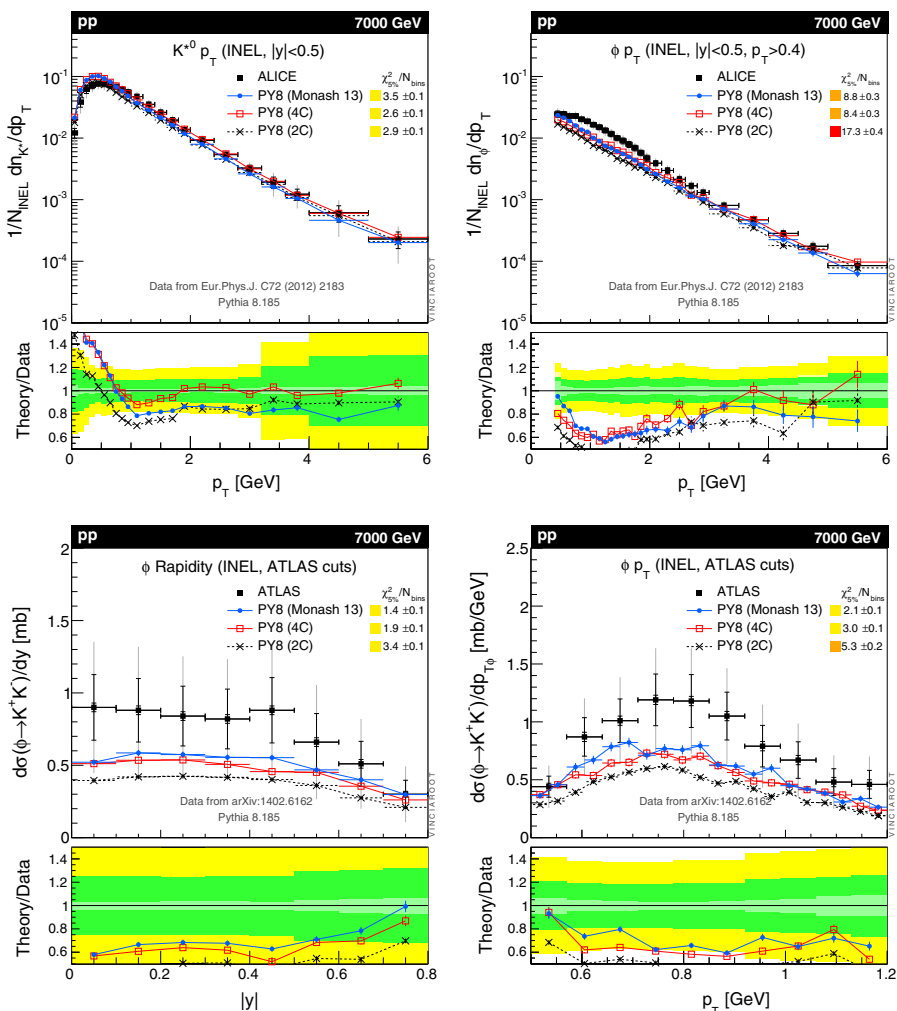
### B.3 Energy scaling of $\sigma_{2 \rightarrow 2}(p_{T\min})$ vs $\sigma_{\text{inel}}$ from 200 GeV to 100 TeV

In Fig. 34, we show the LO QCD  $2 \rightarrow 2$  cross section, integrated above  $p_{T\min}$ , as a function of  $p_{T\min}$ , in  $pp$  collisions



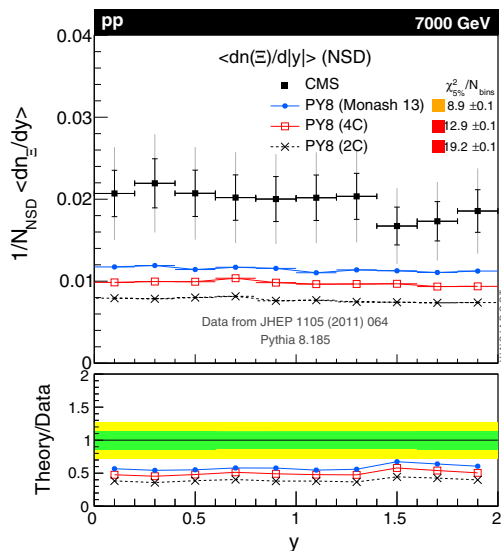
**Fig. 31** Hadronic Z decays at  $\sqrt{s} = 91.2$  GeV.  $p^\pm$  and  $\Xi^\pm$   $x$  spectra

**Fig. 32** *pp* collisions at 7 TeV. Top row  $K^*$  and  $\phi$   $p_\perp$  spectra, compared with ALICE data [142]. Bottom row  $\phi$  rapidity and  $p_\perp$  spectrum, compared with ATLAS data [143]. The ATLAS cuts are  $\phi \rightarrow K^+K^-$ ,  $p_{\perp\phi} \in [0.5, 1.2]$  GeV,  $|y(\phi)| < 0.8$ ,  $p_{\perp K} > 0.23$  GeV,  $|p_K| < 0.8$  GeV

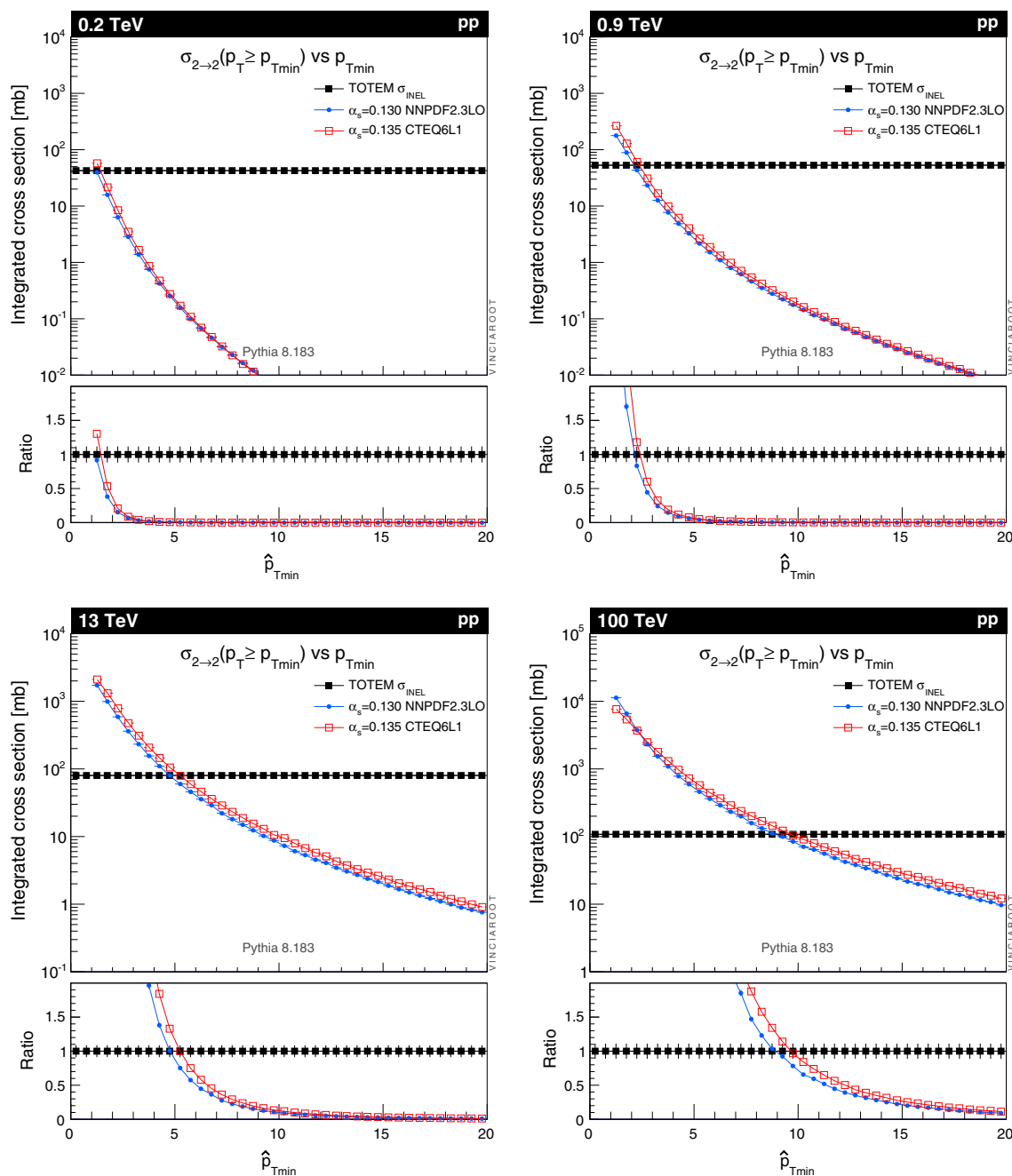


at 4 different CM energies, complementing and expanding on the 8-TeV CM energy shown in the main body of the paper. We compare two different  $\alpha_s$  and PDF choices, corresponding to those made in tunes Monash 13 (blue filled dots) and 4C (red open squares), respectively. As a reference for the total inelastic cross section at each energy, we base ourselves on the best-fit curve in the TOTEM cross-section measurement paper [72], which in turn represents a fit produced by the COMPETE collaboration [144]. Uncertainties are rough conservative estimates based on the plot in the TOTEM paper, but they are in any case too small to significantly affect conclusions about the scale at which the partonic cross section saturates the hadronic one.

We observe that the  $p_{T\min}$  value for which the LO QCD  $2 \rightarrow 2$  partonic cross section formally becomes equal to the total inelastic cross section (strongly suggesting that every event has at least one such mini-jet pair) rises from values around 1–2 GeV at energies  $\sqrt{s} < 1$  TeV, to 5 GeV at  $\sqrt{s} = 13$  TeV, and finally 10 GeV at  $\sqrt{s} = 100$  TeV.



**Fig. 33** *pp* collisions at 7 TeV.  $\Xi^-$  rapidity spectrum, compared with CMS data [99]



**Fig. 34** *pp* collisions at four different CM energies. Integrated QCD  $2 \rightarrow 2$  cross section above  $p_{T\min}$ , as a function of  $p_{T\min}$ . *Top left* 200 GeV; *top right* 900 GeV; *bottom left* 13 TeV; *bottom right* 100 TeV

## References

1. A. Buckley, M. Whalley, HepData reloaded: reinventing the HEP data archive. [1006.0517](https://doi.org/10.1006.0517). <http://hepdata.cedar.ac.uk>
2. A. Buckley, J. Butterworth, L. Lonnblad, D. Grellscheid, H. Hoeth et al., Rivet user manual. *Comput. Phys. Commun.* **184**, 2803–2819 (2013). [1003.0694](https://doi.org/10.1016/j.cpc.2013.06.024)
3. T. Sjöstrand, Colour reconnection and its effects on precise measurements at the LHC. [1310.8073](https://doi.org/10.1007/978-3-319-00021-3_13)
4. T. Sjöstrand, Challenges for QCD theory: some personal reflections. *Phys. Scr. T* **158**, 014002 (2013). [1309.6747](https://doi.org/10.1088/0031-858X/158/1/014002)
5. J.M. Katzy, QCD Monte-Carlo model tunes for the LHC. *Prog. Part. Nucl. Phys.* **73**, 141–187 (2013)
6. T. Sjöstrand, S. Mrenna, P.Z. Skands, PYTHIA 6.4 physics and manual. *JHEP* **0605**, 026 (2006). [hep-ph/0603175](https://doi.org/10.1088/1126-6708/2006/05/026)
7. T. Sjöstrand, S. Mrenna, P.Z. Skands, A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.* **178**, 852–867 (2008). [0710.3820](https://doi.org/10.1016/j.cpc.2007.07.043)
8. P.Z. Skands, Tuning Monte Carlo generators: the Perugia tunes. *Phys. Rev. D* **82**, 074018 (2010). [1005.3457](https://doi.org/10.1103/PhysRevD.82.074018)
9. R. Corke, T. Sjöstrand, Interleaved parton showers and tuning prospects. *JHEP* **1103**, 032 (2011). [1011.1759](https://doi.org/10.1088/1126-6708/2011/03/032)

10. ATLAS Collaboration, Charged particle multiplicities in pp interactions at  $\sqrt{s} = 0.9$  and 7 TeV in a diffractive limited phase-space measured with the ATLAS detector at the LHC and new PYTHIA6 tune. ATLAS-CONF-2010-031, ATLAS-COM-CONF-2010-031
11. ATLAS Collaboration, New ATLAS event generator tunes to 2010 data. ATL-PHYS-PUB-2011-008, ATL-COM-PHYS-2011-329
12. ATLAS Collaboration, ATLAS tunes of PYTHIA 6 and Pythia 8 for MC11. ATL-PHYS-PUB-2011-009, ATL-COM-PHYS-2011-744
13. ATLAS Collaboration, Summary of ATLAS Pythia 8 tunes. ATL-PHYS-PUB-2012-003, ATL-COM-PHYS-2012-738
14. R. Field, The underlying event in hadronic collisions. Ann. Rev. Nucl. Part. Sci. **62**, 453–483 (2012)
15. J. Alcaraz Maestre et al., The SM and NLO multileg and SM MC working groups: summary report. [1203.6803](#)
16. N. Firdous, G. Rudolph, Tuning of PYTHIA6 to minimum bias data. EPJ Web Conf. **60**, 20056 (2013)
17. A. Buckley, H. Hoeth, H. Lacker, H. Schulz, J.E. von Seggern, Systematic event generator tuning for the LHC. Eur. Phys. J. C **65**, 331–357 (2010). [0907.2973](#)
18. The NNPDF Collaboration Collaboration, R.D. Ball et al., Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO. Nucl. Phys. B **855**, 153–221 (2012). [1107.2652](#)
19. NNPDF Collaboration, R.D. Ball et al., Parton distributions with QED corrections. Nucl. Phys. B **877**(2), 290–320 (2013). [1308.0598](#)
20. S. Carrazza, S. Forte, J. Rojo, Parton distributions and event generators. [1311.5887](#)
21. B. Cooper, J. Katzy, M. Mangano, A. Messina, L. Mijovic et al., Importance of a consistent choice of alpha(s) in the matching of AlpGen and Pythia. Eur. Phys. J. C **72**, 2078 (2012). [1109.5295](#)
22. R. Field, C. Group, D. Wilson, The energy dependence of min-bias and the underlying event at CDF. CDF Note 10874 (2012)
23. R. Field, in *Published in the proceedings of the international symposium on multiparticle dynamics*, September 17, 2013, Chicago, Illinois. The Tevatron energy scan: findings and surprises (2013)
24. W. Giele, D. Kosower, P. Skands, Higher-order corrections to timelike jets. Phys. Rev. D **84**, 054003 (2011). [1102.2126](#)
25. A. Karneyeu, L. Mijovic, S. Prestel, P. Skands, MCPLOTS: a particle physics resource based on volunteer computing. [1306.3436](#). <http://mcplots.cern.ch>
26. L3 Collaboration, P. Achard et al., Studies of hadronic event structure in  $e^+e^-$  annihilation from 30-GeV to 209-GeV with the L3 detector. Phys. Rep. **399**, 71–174 (2004). [hep-ex/0406049](#)
27. T. Sjöstrand, P.Z. Skands, Transverse-momentum-ordered showers and interleaved multiple interactions. Eur. Phys. J. C **39**, 129–154 (2005). [hep-ph/0408302](#)
28. S. Catani, B.R. Webber, G. Marchesini, QCD coherent branching and semiinclusive processes at large  $x$ . Nucl. Phys. B **349**, 635–654 (1991)
29. L. Hartgring, E. Laenen, P. Skands, Antenna showers with one-loop matrix elements. JHEP **1310**, 127 (2013). [1303.4974](#)
30. N. Fischer, S. Gieseke, S. Plätzer, P. Skands, Revisiting radiation patterns in  $e^+e^-$  collisions. [1402.3186](#)
31. G. Gustafson, Dual description of a confined color field. Phys. Lett. B **175**, 453 (1986)
32. B. Andersson, G. Gustafson, T. Sjöstrand, A three-dimensional model for quark and gluon jets. Z. Phys. C **6**, 235 (1980)
33. B. Andersson, The Lund model. Camb. Monogr. Part. Phys. Nucl. Phys. Cosm. (1997)
34. A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Höche et al., General-purpose event generators for LHC physics. Phys. Rep. **504**, 145–233 (2011). [1101.2599](#)
35. B. Webber, A QCD model for jet fragmentation including soft gluon interference. Nucl. Phys. B **238**, 492 (1984)
36. B. Andersson, G. Gustafson, T. Sjöstrand, A model for baryon production in quark and gluon jets. Nucl. Phys. B **197**, 45 (1982)
37. P. Skands, Introduction to QCD. Lectures given at TASI 2012. [1207.2389](#)
38. ALEPH Collaboration, R. Barate et al., Studies of quantum chromodynamics with the ALEPH detector. Phys. Rep. **294**, 1–165 (1998)
39. Particle Data Group Collaboration, J. Beringer et al., Review of particle physics (RPP). Phys. Rev. D **86**, 010001 (2012)
40. OPAL Collaboration, K. Ackerstaff et al., Production of  $f(0)(980)$ ,  $f(2)(1270)$  and  $\phi(1020)$  in hadronic Z0 decay. Eur. Phys. J. C **4**, 19–28 (1998). [hep-ex/9802013](#)
41. DELPHI Collaboration, P. Abreu et al., Measurement of inclusive  $K^*(892)$ ,  $\Phi(1020)$  and  $K(2)^*(1430)$  production in hadronic Z decays. Z. Phys. C **73**, 61–72 (1996)
42. SLD Collaboration, K. Abe et al., Production of  $\pi^+$ ,  $K^+$ ,  $K_0$ ,  $K^*(892)$ ,  $\phi$ ,  $p$  and  $\Lambda(1520)$  in hadronic Z0 decays. Phys. Rev. D **59**, 052001 (1999). [hep-ex/9805029](#)
43. OPAL Collaboration, G. Alexander et al., Strange baryon production in hadronic Z0 decays. Z. Phys. C **73**, 569 (1997)
44. L3 Collaboration, M. Acciarri et al., Inclusive  $\Sigma^+$  and  $\Sigma^0$  production in hadronic Z decays. Phys. Lett. B **479**, 79–88 (2000). [hep-ex/0002066](#)
45. DELPHI Collaboration, P. Abreu et al., Inclusive  $\Sigma^-$  and  $\Lambda(1520)$  production in hadronic Z decays. Phys. Lett. B **475**, 429–447 (2000). [hep-ex/0103020](#)
46. OPAL Collaboration, G. Alexander et al.,  $\Delta^{++}$  production in hadronic Z0 decays. Phys. Lett. B **358**, 162–172 (1995)
47. DELPHI Collaboration, P. Abreu et al., Measurement of  $\Delta^{++}$  (1232) production in hadronic Z decays. Phys. Lett. B **361**, 207–220 (1995)
48. DELPHI Collaboration, P. Abreu et al., Strange baryon production in Z hadronic decays. Z. Phys. C **67**, 543–554 (1995)
49. DELPHI Collaboration, W. Adam et al., Production of SIGMA0 and OMEGA- in Z decays. Z. Phys. C **70**, 371–382 (1996)
50. MARK-II Collaboration, G. Abrams, C. Adolphsen, D. Averill, J. Ballam, B. C. Barish et al., Measurements of charged particle inclusive distributions in hadronic decays of the Z boson. Phys. Rev. Lett. **64**, 1334 (1990)
51. DELPHI Collaboration, P. Abreu et al.,  $\pi^+$ ,  $K^+$ ,  $p$  and anti- $p$  production in  $Z^0 \rightarrow q \text{ anti-}q$ ,  $Z^0 \rightarrow b \text{ anti-}b$ ,  $Z^0 \rightarrow u \text{ anti-}u$ ,  $d \text{ anti-}d$ ,  $s \text{ anti-}s$ . Eur. Phys. J. C **5**, 585–620 (1998)
52. OPAL Collaboration, K. Ackerstaff et al., Measurements of flavor dependent fragmentation functions in  $Z^0 \rightarrow q \text{ anti-}q$  events. Eur. Phys. J. C **7**, 369–381 (1999). [hep-ex/9807004](#)
53. L3 Collaboration, B. Adeva et al., Studies of hadronic event structure and comparisons with QCD models at the Z0 resonance. Z. Phys. C **55**, 39–62 (1992)
54. ALEPH Collaboration, D. Buskulic et al., Production of charmed mesons in Z decays. Z. Phys. C **62**, 1–14 (1994)
55. ALEPH Collaboration, R. Barate et al., Study of charm production in Z decays. Eur. Phys. J. C **16**, 597–611 (2000). [hep-ex/9909032](#)
56. OPAL Collaboration, G. Abbiendi et al., Measurement of the production rate of charm quark pairs from gluons in hadronic Z0 decays. Eur. Phys. J. C **13**, 1–13 (2000). [hep-ex/9908001](#)
57. D. Miller, M.H. Seymour, Secondary heavy quark pair production in  $e^+e^-$  annihilation. Phys. Lett. B **435**, 213–220 (1998). [hep-ph/9805414](#)
58. O. Biebel, P. Nason, B. Webber, Jet fragmentation in  $e^+e^-$  annihilation. [hep-ph/0109282](#)
59. ALEPH Collaboration, R. Barate et al., A measurement of the gluon splitting rate into  $b \text{ anti-}b$  pairs in hadronic Z decays. Phys. Lett. B **434**, 437–450 (1998)
60. DELPHI Collaboration, P. Abreu et al., Measurement of the multiplicity of gluons splitting to bottom quark pairs in hadronic Z0 decays. Phys. Lett. B **405**, 202–214 (1997)

61. SLD Collaboration, K. Abe et al., Measurement of the probability for gluon splitting into b anti-b in Z0 decays. [hep-ex/9908028](#)
62. M.G. Bowler,  $e^+e^-$  production of heavy quarks in the string model. *Z. Phys. C* **11**, 169 (1981)
63. DELPHI Collaboration, J. Abdallah et al., A study of the b-quark fragmentation function with the DELPHI detector at LEP I and an averaged distribution obtained at the Z Pole. *Eur. Phys. J. C* **71**, 1557 (2011). [1102.4748](#)
64. SLD Collaboration, K. Abe et al., Measurement of the b quark fragmentation function in Z0 decays. *Phys. Rev. D* **65**, 092006 (2002). [hep-ex/0202031](#)
65. ATLAS Collaboration, G. Aad et al., Measurement of  $D^{*+/-}$  meson production in jets from pp collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector. *Phys. Rev. D* **85**, 052005 (2012). [1112.4432](#)
66. M. Bähr, S. Gieseke, M. Gigg, D. Grellscheid, K. Hamilton et al., Herwig++ physics and manual. *Eur. Phys. J. C* **58**, 639–707 (2008). [0803.0883](#)
67. T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann et al., Event generation with SHERPA 1.1. *JHEP* **0902**, 007 (2009). [0811.4622](#)
68. G. Altarelli, S. Forte, G. Ridolfi, On positivity of parton distributions. *Nucl. Phys. B* **534**, 277–296 (1998). [hep-ph/9806345](#)
69. A. Sherstnev, R.S. Thorne, Parton distributions for LO generators. *Eur. Phys. J. C* **55**, 553–575 (2008). [0711.2473](#)
70. H.-L. Lai et al., Parton distributions for event generators. *JHEP* **04**, 035 (2010). [0910.4183](#)
71. V. Bertone, S. Carrazza, J. Rojo, APFEL: a PDF evolution library with QED corrections. *Comput. Phys. Commun.* **185**, 1647–1668 (2014). [1310.1394](#)
72. TOTEM Collaboration, G. Antchev et al., Luminosity-independent measurement of the proton-proton total cross section at  $\sqrt{s} = 8$  TeV. *Phys. Rev. Lett.* **111**(1), 012001 (2013)
73. A. Donnachie, P. Landshoff, Total cross-sections. *Phys. Lett. B* **296**, 227–232 (1992). [hep-ph/9209205](#)
74. G.A. Schuler, T. Sjöstrand, Towards a complete description of high-energy photoproduction. *Nucl. Phys. B* **407**, 539–605 (1993)
75. A. Donnachie, P. Landshoff,  $pp$  and  $\bar{p}p$  total cross sections and elastic scattering. *Phys. Lett. B* **727**, 500–505 (2013). [1309.1292](#)
76. T. Sjöstrand, M. van Zijl, A multiple interaction model for the event structure in hadron collisions. *Phys. Rev. D* **36**, 2019 (1987)
77. T. Sjöstrand, P.Z. Skands, Multiple interactions and the structure of beam remnants. *JHEP* **0403**, 053 (2004). [hep-ph/0402078](#)
78. M. Bähr, J.M. Butterworth, M.H. Seymour, The underlying event and the total cross section from Tevatron to the LHC. *JHEP* **0901**, 065 (2009). [0806.2949](#)
79. G. Gustafson, U. Pettersson, Dipole formulation of QCD cascades. *Nucl. Phys. B* **306**, 746 (1988)
80. W.T. Giele, D.A. Kosower, P.Z. Skands, A simple shower and matching algorithm. *Phys. Rev. D* **78**, 014026 (2008). [0707.3652](#)
81. M. Ritzmann, D. Kosower, P. Skands, Antenna showers with hadronic initial states. *Phys. Lett. B* **718**, 1345–1350 (2013). [1210.6345](#)
82. ATLAS Collaboration, G. Aad et al., Measurement of the transverse momentum distribution of Z/gamma\* bosons in proton-proton collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector. *Phys. Lett. B* **705**, 415–434 (2011). [1107.2381](#)
83. A. Buckley, G. Hesketh, F. Siegert, P. Skands, M. Vesterinen, T. Wyatt, in *Tools and Monte Carlo working group: summary report*, Les Houches, France. Effect of QED FSR on measurements of  $Z/\gamma^*$  and  $W$  leptonic final states at hadron colliders (2009). [1003.1643](#)
84. CDF Collaboration, T. Affolder et al., The transverse momentum and total cross section of  $e^+e^-$  pairs in the Z boson region from  $p\bar{p}$  collisions at  $\sqrt{s} = 1.8$  TeV. *Phys. Rev. Lett.* **84**, 845–850 (2000). [hep-ex/0001021](#)
85. G. Miu, T. Sjöstrand,  $W$  production in an improved parton shower approach. *Phys. Lett. B* **449**, 313–320 (1999). [hep-ph/9812455](#)
86. S. Frixione, P. Nason, C. Oleari, Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP* **0711**, 070 (2007). [0709.2092](#)
87. K. Hamilton, P. Nason, C. Oleari, G. Zanderighi, Merging H/W/Z + 0 and 1 jet at NLO with no merging scale: a path to parton shower + NNLO matching. *JHEP* **1305**, 082 (2013). [1212.4504](#)
88. CMS Collaboration, V. Khachatryan et al., Observation of long-range near-side angular correlations in proton-proton collisions at the LHC. *JHEP* **1009**, 091 (2010). [1009.4122](#)
89. R. Corke, T. Sjöstrand, Multiparton interactions and rescattering. *JHEP* **1001**, 035 (2010). [0911.1909](#)
90. A. Ortiz, P. Christiansen, E. Cuautle, I. Maldonado, G. Paic, Color reconnection and flow-like patterns in pp collisions. *Phys. Rev. Lett.* **111**, 042001 (2013). [1303.6326](#)
91. ATLAS Collaboration, G. Aad et al., Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC. *New J. Phys.* **13**, 053033 (2011). [1012.5104](#)
92. CMS Collaboration, V. Khachatryan et al., Transverse-momentum and pseudorapidity distributions of charged hadrons in  $pp$  collisions at  $\sqrt{s} = 7$  TeV. *Phys. Rev. Lett.* **105**, 022002 (2010). [1005.3299](#)
93. CMS Collaboration, S. Chatrchyan et al., Measurement of energy flow at large pseudorapidities in  $pp$  collisions at  $\sqrt{s} = 0.9$  and 7 TeV. *JHEP* **1111**, 148 (2011). [1110.0211](#)
94. TOTEM Collaboration, G. Antchev et al., Measurement of the forward charged particle pseudorapidity density in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the TOTEM experiment. *Europhys. Lett.* **98**, 31002 (2012). [1205.4105](#)
95. P.Z. Skands, D. Wicke, Non-perturbative QCD effects and the top mass at the Tevatron. *Eur. Phys. J. C* **52**, 133–140 (2007). [hep-ph/0703081](#)
96. CDF Collaboration, T. Aaltonen et al., Measurement of particle production and inclusive differential cross sections in p anti-p collisions at  $s^{*1/2} = 1.96$ -TeV. *Phys. Rev. D* **79**, 112005 (2009). [0904.1098](#)
97. H. Schulz, P. Skands, Energy scaling of minimum-bias tunes. *Eur. Phys. J. C* **71**, 1644 (2011). [1103.3649](#)
98. ATLAS Collaboration, G. Aad et al., Measurement of underlying event characteristics using charged particles in pp collisions at  $\sqrt{s} = 900$  GeV and 7 TeV with the ATLAS detector. *Phys. Rev. D* **83**, 112001 (2011). [1012.0791](#)
99. CMS Collaboration, V. Khachatryan et al., Strange particle production in  $pp$  collisions at  $\sqrt{s} = 0.9$  and 7 TeV. *JHEP* **1105**, 064 (2011). [1102.4282](#)
100. TASSO Collaboration, W. Braunschweig et al., Charged multiplicity distributions and correlations in  $e^+e^-$  annihilation at PETRA energies. *Z. Phys. C* **45**, 193 (1989)
101. M. Derrick, K. Gan, P. Kooijman, J. Loos, B. Musgrave et al., Study of quark fragmentation in  $e^+e^-$  annihilation at 29-GeV: charged particle multiplicity and single particle rapidity distributions. *Phys. Rev. D* **34**, 3304 (1986)
102. TOPAZ Collaboration, K. Nakabayashi et al., Charged particle multiplicities of quark and gluon jets in  $e^+e^-$  annihilation at TRISTAN. *Phys. Lett. B* **413**, 447–452 (1997)
103. ALEPH Collaboration, D. Buskulic et al., Studies of QCD in  $e^+e^- \rightarrow$  hadrons at  $E(\text{cm}) = 130$ -GeV and 136-GeV. *Z. Phys. C* **73**, 409–420 (1997)
104. OPAL Collaboration, G. Alexander et al., QCD studies with  $e^+e^-$  annihilation data at 130-GeV and 136-GeV. *Z. Phys. C* **72**, 191–206 (1996)
105. DELPHI Collaboration, P. Abreu et al., Charged particle multiplicity in  $e^+e^- \rightarrow q$  anti- $q$  events at 161-GeV and 172-GeV and from the decay of the W boson. *Phys. Lett. B* **416**, 233–246 (1998)

106. OPAL Collaboration, K. Ackerstaff et al., QCD studies with  $e^+e^-$  annihilation data at 161-GeV. *Z. Phys. C* **75**, 193–207 (1997)
107. OPAL Collaboration, G. Abbiendi et al., QCD studies with  $e^+e^-$  annihilation data at 172-GeV–189-GeV. *Eur. Phys. J. C* **16**, 185–210 (2000). [hep-ex/0002012](#)
108. DELPHI Collaboration, P. Abreu et al., Charged and identified particles in the hadronic decay of W bosons and in  $e^+e^- \rightarrow q\bar{q}$  from 130-GeV to 200-GeV. *Eur. Phys. J. C* **18**, 203–228 (2000). [hep-ex/0103031](#)
109. CMS Collaboration, V. Khachatryan et al., Transverse momentum and pseudorapidity distributions of charged hadrons in pp collisions at  $\sqrt{s} = 0.9$  and 2.36 TeV. *JHEP* **1002**, 041 (2010). [1002.0621](#)
110. UA5 Collaboration, G. Alner et al., Scaling of pseudorapidity distributions at c.m. energies up to 0.9-TeV. *Z. Phys. C* **33**, 1–6 (1986)
111. UA5 Collaboration, R. Ansorge et al., Charged particle multiplicity distributions at 200-GeV and 900-GeV center-of-mass energy. *Z. Phys. C* **43**, 357 (1989)
112. D. Bourilkov, R.C. Group, M.R. Whalley, LHAPDF: PDF use from the Tevatron to the LHC. [hep-ph/0605240](#)
113. P. Skands, B. Webber, J. Winter, QCD coherence and the top quark asymmetry. *JHEP* **1207**, 151 (2012). [1205.1466](#)
114. J. Winter, P.Z. Skands, B.R. Webber, Monte Carlo event generators and the top quark forward–backward asymmetry. *EPJ Web Conf.* **49**, 17001 (2013). [1302.3164](#)
115. P. Richardson, D. Winn, Investigation of Monte Carlo uncertainties on Higgs boson searches using jet substructure. *Eur. Phys. J. C* **72**, 2178 (2012). [1207.0380](#)
116. S. Navin, Diffraction in pythia. [1005.3894](#)
117. R. Ciesielski, K. Goulianos, MBR Monte Carlo simulation in PYTHIA8. *PoS ICHEP 2012*, 301 (2013). [1205.1446](#)
118. UA5 Collaboration, R. Ansorge et al., Charged particle correlations in  $\bar{P}P$  collisions at c.m. energies of 200-GeV, 546-GeV and 900-GeV. *Z. Phys. C* **37**, 191–213 (1988)
119. K. Wraight, P. Skands, Forward–backward correlations and event shapes as probes of minimum-bias event properties. *Eur. Phys. J. C* **71**, 1628 (2011). [1101.5215](#)
120. C. Søgaard, Measurement of forward–backward charged particle correlations with ALICE. PhD thesis, Copenhagen University, 2012
121. E. Sicking, Multiplicity dependence of two-particle angular correlations in proton–proton collisions measured with ALICE at the LHC. PhD thesis, Münster University. CERN-THESIS-2012-210, 2012
122. ALICE Collaboration, G. Feofilov et al., Forward–backward multiplicity correlations in pp collisions in ALICE at 0.9, 2.76 and 7 TeV. *PoS Baldin-ISHEPP XXI*, 075 (2012)
123. ALICE Collaboration, B. Abelev et al., Multiplicity dependence of two-particle azimuthal correlations in pp collisions at the LHC. *JHEP* **1309**, 049 (2013). [1307.1249](#)
124. D. Lange, The EvtGen particle decay simulation package. *Nucl. Instrum. Methods A* **462**, 152–155 (2001)
125. R. Corke, T. Sjöstrand, Improved parton showers at large transverse momenta. *Eur. Phys. J. C* **69**, 1–18 (2010). [1003.2384](#)
126. S. Höche, F. Krauss, M. Schönherr, Uncertainties in MEPS@NLO calculations of h+jets. [1401.7971](#)
127. M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A.D. Polosa, ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP* **0307**, 001 (2003). [hep-ph/0206293](#)
128. J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, T. Stelzer, MadGraph 5: going beyond. *JHEP* **1106**, 128 (2011). [1106.0522](#)
129. V. Hirschi, R. Frederix, S. Frixione, M.V. Garzelli, F. Maltoni et al., Automation of one-loop QCD corrections. *JHEP* **1105**, 044 (2011). [1103.0621](#)
130. S. Alioli, P. Nason, C. Oleari, E. Re, A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP* **1006**, 043 (2010). [1002.2581](#)
131. J. Alwall, A. Ballestrero, P. Bartalini, S. Belov, E. Boos et al., A standard format for Les Houches event files. *Comput. Phys. Commun.* **176**, 300–304 (2007). [hep-ph/0609017](#)
132. S. Catani, F. Krauss, R. Kuhn, B.R. Webber, QCD matrix elements + parton showers. *JHEP* **11**, 063 (2001). [hep-ph/0109231](#)
133. L. Lönnblad, Correcting the color dipole cascade model with fixed order matrix elements. *JHEP* **0205**, 046 (2002). [hep-ph/0112284](#)
134. N. Lavesson, L. Lönnblad, W+jets matrix elements and the dipole cascade. *JHEP* **0507**, 054 (2005). [hep-ph/0503293](#)
135. M.L. Mangano, M. Moretti, F. Piccinini, M. Treccani, Matching matrix elements and shower evolution for top-quark production in hadronic collisions. *JHEP* **0701**, 013 (2007). [hep-ph/0611129](#)
136. S. Mrenna, P. Richardson, Matching matrix elements and parton showers with HERWIG and PYTHIA. *JHEP* **0405**, 040 (2004). [hep-ph/0312274](#)
137. L. Lönnblad, S. Prestel, Unitarising matrix element + parton shower merging. *JHEP* **1302**, 094 (2013). [1211.4827](#)
138. N. Lavesson, L. Lönnblad, Extending CKKW-merging to one-loop matrix elements. *JHEP* **0812**, 070 (2008). [0811.2912](#)
139. L. Lönnblad, S. Prestel, Merging multi-leg NLO matrix elements with parton showers. *JHEP* **1303**, 166 (2013). [1211.7278](#)
140. T. Pierog, I. Karpenko, J. Katzy, E. Yatsenko, K. Werner, EPOS LHC: test of collective hadronization with LHC data. [1306.0121](#)
141. T. Sjöstrand, P.Z. Skands, Baryon number violation and string topologies. *Nucl. Phys. B* **659**, 243 (2003). [hep-ph/0212264](#)
142. ALICE Collaboration, B. Abelev et al., Production of  $K^*(892)^0$  and  $\phi(1020)$  in pp collisions at  $\sqrt{s} = 7$  TeV. *Eur. Phys. J. C* **72**, 2183 (2012). [1208.5717](#)
143. ATLAS Collaboration, G. Aad et al., The differential production cross section of the  $\phi(1020)$  meson in  $\sqrt{s} = 7$  TeV pp collisions measured with the ATLAS detector. *Eur. Phys. J. C* **74**, 2895 (2014). [1402.6162](#)
144. COMPETE Collaboration, J. Cudell et al., Benchmarks for the forward observables at RHIC, the Tevatron Run II and the LHC. *Phys. Rev. Lett.* **89**, 201801 (2002). [hep-ph/0206172](#)