

Thermodynamical String Fragmentation

Nadine Fischer^{a,b} and Torbjörn Sjöstrand^a

^a *Theoretical Particle Physics, Department of Astronomy and Theoretical Physics,
Lund University, SE-223 62 Lund, Sweden*

^b *School of Physics and Astronomy, Monash University,
Clayton VIC-3800, Australia*

Abstract

The observation of heavy-ion-like behaviour in pp collisions at the LHC suggests that more physics mechanisms are at play than traditionally assumed. The introduction e.g. of quark-gluon plasma or colour rope formation can describe several of the observations, but as of yet there is no established paradigm. In this article we study a few possible modifications to the PYTHIA event generator, which describes a wealth of data but fails for a number of recent observations. Firstly, we present a new model for generating the transverse momentum of hadrons during the string fragmentation process, inspired by thermodynamics, where heavier hadrons naturally are suppressed in rate but obtain a higher average transverse momentum. Secondly, close-packing of strings is taken into account by making the temperature or string tension environment-dependent. Thirdly, a simple model for hadron rescattering is added. The effect of these modifications is studied, individually and taken together, and compared with data mainly from the LHC. While some improvements can be noted, it turns out to be nontrivial to obtain effects as big as required, and further work is called for.

1 Introduction

QCD, the theory of strong interactions, is at the origin of a wide range of phenomena. In one extreme, progress on high-energy perturbative calculations offers an increasingly precise and successful description of hard processes, as a large community is steadily improving calculational techniques. NLO calculations, once rare, are now standard, NNLO is getting there, and even NNNLO is starting to appear (see e.g. [1] and references therein). In another extreme, the nonperturbative aspects of low-energy interactions are less well understood. Lattice QCD can be used to calculate static hadron properties, but not (yet?) dynamical processes. Specifically, the description of hadronization, the step whereby partons turn into hadrons in high-energy collisions, cannot be derived directly from the QCD Lagrangian within any currently known formalism. Instead string [2] and cluster [3–5] models, developed in the early eighties, have been used almost unchanged from PETRA/LEP e^+e^- events to Sp \bar{p} S/Tevatron/LHC $p\bar{p}/pp$ ones — the assumed “jet universality”. Differences have been attributed to the quite disparate parton-level configurations that undergo hadronization: while e^+e^- involves only hard process and final-state radiation (FSR), pp adds aspects such as initial-state radiation (ISR), multiparton interactions (MPIs), beam remnants and colour reconnection (CR).

Cracks have started to appear in this picture as new LHC data have been presented. Specifically, several studies have shown how high-multiplicity pp events have properties similar to those observed in heavy-ion AA collisions. Some observations may have an explanation within the current framework, e.g. CR may give some flow-like patterns [6], but others do not. An early example was the discovery of “the ridge”, an enhanced particle production around the azimuthal angle of a trigger jet, stretching away in (pseudo)rapidity [7–9]. A more recent example is the smoothly increasing fraction of strange baryon production with increasing charged multiplicity, a trend that lines up with pA data before levelling out at the AA results [10]. Conventional wisdom holds that the formation of a quark–gluon plasma (QGP) requires a larger volume and longer time for thermalization than pp or pA systems can offer, so such trends are unexpected, see e.g. [11].

It is therefore time to rethink the picture of hadronization in high-energy and high-multiplicity collisions. One possible approach is to imagine that a QGP is at least partly formed in pp collisions, such that individual colour fields (strings) cease to exist. Such a behaviour is already implemented in the EPOS model [12]. Another is to imagine that strings survive as a vehicle e.g. of short-range flavour correlations, but that their properties are modified. Colour ropes [13–15] is one such example, wherein several colour-triplet strings combine to a higher colour-representation field. A detailed implementation of rope dynamics is found in the DIPSY program [16]. Both EPOS and DIPSY qualitatively describe several of the new key features, such as the increasing rate of strangeness production at higher multiplicities.

With the studies described in this article we want to add to the set of alternative models that can be used to compare with data. At best it may offer some new insights, at worst it will act as a straw man model. Firstly, rather than the particle-mass-independent Gaussian p_\perp spectrum assumed in the standard string model, it introduces an exponential p_\perp dependence, $\exp(-p_\perp/T)$. This is split among possible flavours according to hadronic m_\perp , $\exp(-m_\perp/T)$. Such p_\perp and m_\perp shapes were used to describe early pp data, e.g. at the ISR [17], and has some foundation in the Hagedorn temperature [18–20] and in related [21]

ideas. (Later powerlike p_{\perp} ansätze [20, 22, 23] or two-component exponential + powerlike ones [24] can be viewed as a consequence of perturbative jet production, and is in our framework generated as such, in an earlier stage than the nonperturbative hadronization.) Secondly, it assumes that the close-packing of several strings leads to an increased effective temperature and thereby both a changed particle composition and changed p_{\perp} spectra. In spirit this is close to the rope model, but it does not have to assume that the individual strings either fuse or melt away. Thirdly, if the fragmenting strings are close-packed this also implies the initial formation of a dense hadronic gas, wherein rescattering may lead to collective-flow effects. Such effects are simulated in a crude first approximation.

The impact of these mechanisms on experimental distributions is studied, in order to quantify their significance. As a prime example, consider the $\langle p_{\perp} \rangle(n_{\text{ch}})$ distribution, with a characteristic rising trend that has been proposed as a signal for colour reconnection [25]. Alternative interpretations are now offered in terms of close-packing of strings and/or hadrons, and these are presented and compared with data individually. At the end of the day, we should expect the “true” nature of high-multiplicity pp collisions to contain many contributing mechanisms, however. To be more more specific, in quantum mechanics any process that is not explicitly forbidden by some selection rule is bound to occur, the question is only with what rate. The final task therefore is try to constrain the relative importance of the mechanisms, not to prove a specific one “right” or “wrong”.

The new model components are implemented as options in the standard PYTHIA event generator [26, 27], which makes it easily accessible for further experimental tests. They should be viewed as a first iteration. Should they prove useful there is room for further improvements, as we will indicate.

The article is organized as follows. Sec. 2 outlines relevant features of the existing Lund string model and introduces key observables, with emphasis on those new ones that are not well described by the current PYTHIA generator. Sec. 3 introduces the alternative approaches explored in this article, and presents some first toy studies for simplified string topologies. Comparisons with data are presented in sec. 4, highlighting what seems to work where and what not. Finally sec. 5 contains a summary and outlook.

2 Existing Models and Data

2.1 The Lund string model

The Lund string fragmentation model [2] is very successful in many respects, but more so for the overall longitudinal fragmentation structure than for its description of the particle composition.

The central assumption in the string model is that of linear confinement, $V(r) = \kappa r$, with a string tension $\kappa \approx 1$ GeV/fm. The word “string” should here not be taken literally; the physical object is a kind of flux tube stretched between the endpoints, with a typical transverse size of the order of the proton one, $r_p \sim 0.7$ fm. The one-dimensional “mathematical” string should then be viewed as a description of the location of the center of the flux tube. By analogy with superconductivity the tube could be viewed as a vortex line like in a type II superconductor, alternatively as an elongated bag in a type I one.

In the case of a simple stable back-to-back $q\bar{q}$ system, with $m_q = p_{\perp q} = 0$, quarks move

with the speed of light in “yo-yo”-mode oscillations, as energy moves between being stored in the endpoint quarks and in the intermediate string. If creation of new $q\bar{q}$ pairs is allowed the original system can break up into smaller ones, each a colour singlet in its own right. Denoting the original pair $q_0\bar{q}_0$, and ordering the new pairs $q_i\bar{q}_i$, $1 \leq i \leq n-1$ from the quark end, results in the production of n hadrons $q_0\bar{q}_1, q_1\bar{q}_2, \dots, q_{n-1}\bar{q}_0$.

Aligning the x axis with the string axis, the breakup vertices are characterized by their location (t_i, x_i) . These vertices have a spacelike separation, and so have no unique time ordering. (Except for the original $(t_0, x_0) = (0, 0)$ of course. But here it is actually the turning points of the q_0 and \bar{q}_0 that define the vertices in eq. (1) below, and then spacelike separation is restored.) Two adjacent ones are correlated by the constraint that the hadron produced should have the correct mass m_i :

$$\kappa^2((x_i - x_{i-1})^2 - (t_i - t_{i-1})^2) = m_i^2. \quad (1)$$

If the vertices are assigned from the quark end, say, each new vertex therefore corresponds to one degree of freedom, which should be selected according to some probability function. Imposing consistency constraints, mainly that results should be the same (on the average) if fragmentation is instead considered from the antiquark end, gives the solution [2]

$$f(z) \propto \frac{1}{z} (1-z)^a \exp(-bm^2/z), \quad (2)$$

with a and b two free parameters, and where $m^2 \rightarrow m_{\perp}^2$ once transverse momentum is introduced. Here z is the fraction of available lightcone momentum $E + p_x$ taken by a hadron, with the remainder $1-z$ retained by the string for subsequent particle production.

This ansatz leads to vertices having an equilibrium distribution (after having taken a few steps away from the endpoints)

$$P(\Gamma) \propto \Gamma^a \exp(-b\Gamma), \quad \Gamma = (\kappa\tau)^2 = \kappa^2(t^2 - x^2), \quad (3)$$

with the same a and b as above. (For the special case $a = 0$ this result agrees with the Artru-Mennessier model [28], which is based on constant decay probability per string area $dt dx$, without any mass constraint.) The associated probability for producing n particles can be written as [29]

$$dP_n \propto \left[\prod_{i=1}^n N d^2 p_i \delta(p_i^2 - m_i^2) \right] \delta^{(2)} \left(\sum_i p_i - p_{\text{tot}} \right) \exp(-b\kappa^2 A_{\text{tot}}), \quad (4)$$

where A_{tot} is the total space-time area under the breakup vertices. The relation between dP_n and dP_{n-1} (at a reduced c.m. energy) is then given by the fragmentation function eq. (2), where it is easy to show that the exponentials match, and somewhat less trivial that a larger N (i.e. larger weight for higher multiplicities) corresponds to a larger a (i.e. less momentum taken away in each step).

The simple $q\bar{q}$ fragmentation picture can be extended to $q\bar{q}g$ topologies if the gluon is viewed as having separate colour and anticolour indices, as in the $N_C \rightarrow \infty$ limit [30]. Then one string piece is stretched between the quark and the gluon, and another between the gluon and the antiquark. The absence of a string piece stretched directly between the quark and antiquark leads to predicted asymmetries in the particle production [31] that rapidly

were observed experimentally [32]. In general, a string can stretch from a quark end via a number of intermediate gluons to an antiquark end. Technically the motion and fragmentation of such a string system can become rather complicated [33], but the fragmentation can be described without the introduction of any new principles or parameters. This is the most powerful and beautiful aspect of the string fragmentation framework. Note that the leading hadron in a gluon jet can take momentum from both the string pieces that attaches it to colour-adjacent partons. This is unlike cluster models, where gluons are forced to branch into $q\bar{q}$ pairs, such that smaller colour singlets are formed rather than one single string winding its way between the partons. The string model is easily extended to closed gluon loops and, with rather more effort [34], to junction topologies, where three string pieces come together in a single vertex.

We now turn to the breakup mechanism. If a $q\bar{q}$ pair is massless and has no transverse momentum it can be produced on-shell, in a single vertex, and then the q and \bar{q} can move apart, splitting the string into two in the process. But if the q (and \bar{q}) transverse mass $m_{\perp q} = \sqrt{m_q^2 + p_{\perp q}^2} > 0$ this is no longer possible. By local flavour conservation the $q\bar{q}$ pair is still produced at a common vertex, but as virtual particles that each needs to tunnel out a distance $d = m_{\perp}/\kappa$. Using the WKB approximation [2] to calculate the tunneling probability for the pair gives a factor

$$\exp(-\pi m_{\perp q}^2/\kappa) = \exp(-\pi m_q^2/\kappa) \exp(-\pi p_{\perp q}^2/\kappa) \quad , \quad (5)$$

where the Gaussian answer allows a convenient separation of the m and p_{\perp} dependencies (with implicit phase space d^2p_{\perp}).

The latter is implemented by giving the q and \bar{q} opposite and compensating p_{\perp} kicks, with $\langle p_{\perp q}^2 \rangle = \kappa/\pi = \sigma^2 \approx (0.25 \text{ GeV})^2$. A hadron receives its p_{\perp} as the vector sum of its q and \bar{q} constituent kicks, and thus $\langle p_{\perp \text{had}}^2 \rangle = 2\sigma^2$. Empirically the tuned σ value comes out larger than this, actually closer to $\sigma = 0.35 \text{ GeV}$. This implies that almost half of the p_{\perp}^2 kick is coming from other sources than tunneling. One source could be soft gluon radiation below the perturbative (parton shower) cutoff, where α_s becomes so big that perturbation theory breaks down [35]. Effectively radiation near the perturbative/nonperturbative border is thus shoved into an artificially enhanced tunneling answer, with the further assumption that the Gaussian shape and the p_{\perp} balancing inside each new $q\bar{q}$ pair still holds.

Uncertainties also arise in the interpretation of the mass suppression factor of eq. (5): what quark masses to use? If current quark masses then the u and d ones are negligible while the s is below 0.2 GeV , predicting less strangeness suppression than observed, while with constituent masses $m_u \approx m_d \approx 0.33 \text{ GeV}$ and $m_s \approx 0.51 \text{ GeV}$ [36] too much suppression is predicted. Intermediate masses and suppression factors closer to data can be motivated e.g. by noting that an expanding string corresponds to confinement in the two transverse dimensions but not in the longitudinal one. In the end, however, the s/u suppression is viewed as an empirical number to be tuned to data. Whichever values are used, c and b quark tunneling production is strongly suppressed, so this mechanism can be totally neglected relative to the perturbative ones.

Considering only mesons in radial and rotational ground states, i.e. only the pseudoscalar and vector multiplets, naive spin counting predicts relative rates $1 : 3$, whereas data prefers values closer to $1 : 1$, at least for $\pi : \rho$. It is possible to explain a suppression of the vector mesons based on the difference in the hadronic wave functions, from the spin-spin

interaction term [2], but the amount has to be tuned to data. And further brute-force suppression factors are needed specifically for the η and η' mesons, which have “unnaturally” large masses owing to the $U(1)$ anomaly.

Baryon production can be introduced by allowing diquark–antidiquark breakups of the string [37], to be viewed as occurring in two consecutive $q\bar{q}$ creation steps [38]. A baryon and the matching antibaryon would normally be nearest neighbours along the string, but the “popcorn mechanism” also allows one (or more) mesons to be produced in between. Diquark masses can be used to derive approximate suppressions, but again free parameters are used, for qq/q , sq/qq , qq_1/qq_0 and others. Unfortunately the tuned values do not always match so well with the tunneling-formula expectations.

In total $\mathcal{O}(20)$ parameters are used to describe the outcome of the string/tunneling mechanism for particle production. Notable is that the particle masses do not enter explicitly in these considerations. This is unlike cluster models, e.g., where hadron masses occur in the phase space available for different cluster decay channels. A fair overall description of the particle composition is then obtained with very few parameters [39,40]. Note that while most fragmentation parameters in HERWIG++ exist in different copies for light (u, d, s), c, and b quarks, the ones for heavy quarks have either been set equal to the values of those for light quarks [39] or have not been included in further tuning processes [40].

The hadron masses can be explicitly introduced into the Lund framework by assuming that the integral $\int_0^1 f(z) dz$, with $f(z)$ given by eq. (2), provides the relative normalization of possible particle states. This concept has been developed successfully within the UCLA model [41,42], in that particle rates come out quite reasonably with minimal further assumptions. There are some other issues with this approach, however, and we do not pursue it further here.

2.2 Key data

An immense number of studies have been published based on hadron collider data, and it is not the intention here to survey all of that. Instead we here bring up some of the key data and distributions that have prompted us to this study. Several of them will be shown repeatedly in the following. We note that all histograms we will present in this article are produced by utilizing RIVET [43].

The list of key observables includes:

- The change of flavour composition with event multiplicity. Specifically, high-multiplicity events have a higher fraction of heavier particles, meaning particles with a higher strangeness content [10]. PYTHIA contains no mechanism to generate such a behaviour. On the contrary, within a single fixed-energy string a higher multiplicity means more lighter particles, for phase space reasons. In pp collisions a higher multiplicity is predominantly obtained by more MPIs, however, so the composition stays rather constant.
- The average transverse momentum $\langle p_\perp \rangle$ is larger for heavier particles, both at RHIC [44] and LHC [45]. This is a behaviour that is present also in PYTHIA, and comes about quite naturally e.g. by lighter particles more often being decay products, with characteristic $\langle p_\perp \rangle$ values smaller than the primary particles in the string fragmentation. The mass dependence is underestimated, however. That is, π^\pm obtains a

too large $\langle p_{\perp} \rangle$ in PYTHIA and baryons a too small one. Recently $\langle p_{\perp} \rangle$ has also been presented as a function of n_{ch} , inclusive [46] and for different hadron species [47], providing a more differential information on this mismatch. In fig. 1 we show these observables and compare default PYTHIA with data, with the above expected conclusions. Note that the data in fig. 1 of [47] is not (yet) publicly available. To obtain an estimate of the data that is comparable to MC predictions we used an estimate of the logarithmic fits shown in fig. 1 of [47] and used n_{ch} values on the x axis rather than $\langle dn_{\text{ch}}/d\eta \rangle_{|\eta|<0.5}$.

- The charged particle p_{\perp} spectrum is not correctly modelled at low p_{\perp} scales, with PYTHIA producing too few particles at very low values [46, 48, 49]. Often tunes then compensate by producing a bit too many at intermediate p_{\perp} scales. The issue shows up e.g. in minimum-bias $dn_{\text{ch}}/d\eta$ distributions, where it is not possible to obtain a good description for data analyzed with $p_{\perp} > 0.1$ GeV and $p_{\perp} > 0.5$ GeV simultaneously.
- In the p_{\perp} spectra for identified particles [50] it turns out that the deficit at low p_{\perp} is from too little π^{\pm} production. This is not unexpected, given the previous two points, but stresses the need to revise the mass dependence of p_{\perp} spectra.
- The Λ/K p_{\perp} spectrum ratio, measured by CMS [51], where PYTHIA is not able to reproduce the peak at ~ 2.5 GeV completely and overshoots the distribution for large- p_{\perp} values.
- The observation of a ridge in pp collisions was one of the major surprises in the 7 TeV data [7], and has been reconfirmed in the 13 TeV one [8, 9]. The ridge is most clearly visible at the very highest multiplicities, but more careful analyses hints the effect is there, to a smaller extent, also at lower multiplicities. Like in heavy-ion collisions one may also seek a description in terms of correlation functions, $C(\Delta\phi) \propto 1 + \sum_{n \geq 2} v_n \cos(n\Delta\phi)$, notably the v_2 coefficient, with a similar message. These phenomena are not at all described by PYTHIA: there is no mechanism that produces a ridge and, once the effects of back-to-back jet production have been subtracted, also no rise of v_2 .

There are also some other reference distributions that have to be checked. These are ones that already are reasonably well described, but that inevitably would be affected by the introduction of new mechanisms.

- The charged particle multiplicity distribution $P(n_{\text{ch}})$ is sensitive to all mechanisms in minimum-bias physics, but especially the MPI and CR modelling. A mismatch in $\langle n_{\text{ch}} \rangle$ is most easily compensated by modifying the $p_{\perp 0}$ scale of the MPI description. This parameter is used to tame the dp_{\perp}^2/p_{\perp}^4 divergence of the QCD cross section to a finite $dp_{\perp}^2/(p_{\perp 0}^2 + p_{\perp}^2)^2$ shape. It can be viewed as the the inverse of the typical colour screening distance inside the proton. A mismatch in the width of the n_{ch} distribution can be compensated by a modified shape of the b impact-parameter distribution of the two colliding protons. Specifically, a distribution more sharply peaked at $b = 0$ gives a longer tail towards high multiplicities.
- An $\langle p_{\perp} \rangle$ increasing with n_{ch} was noted already by UA1 [52], and has remained at higher energies [46, 47]. It offers a key argument for introducing CR in pp/p \bar{p} collisions,

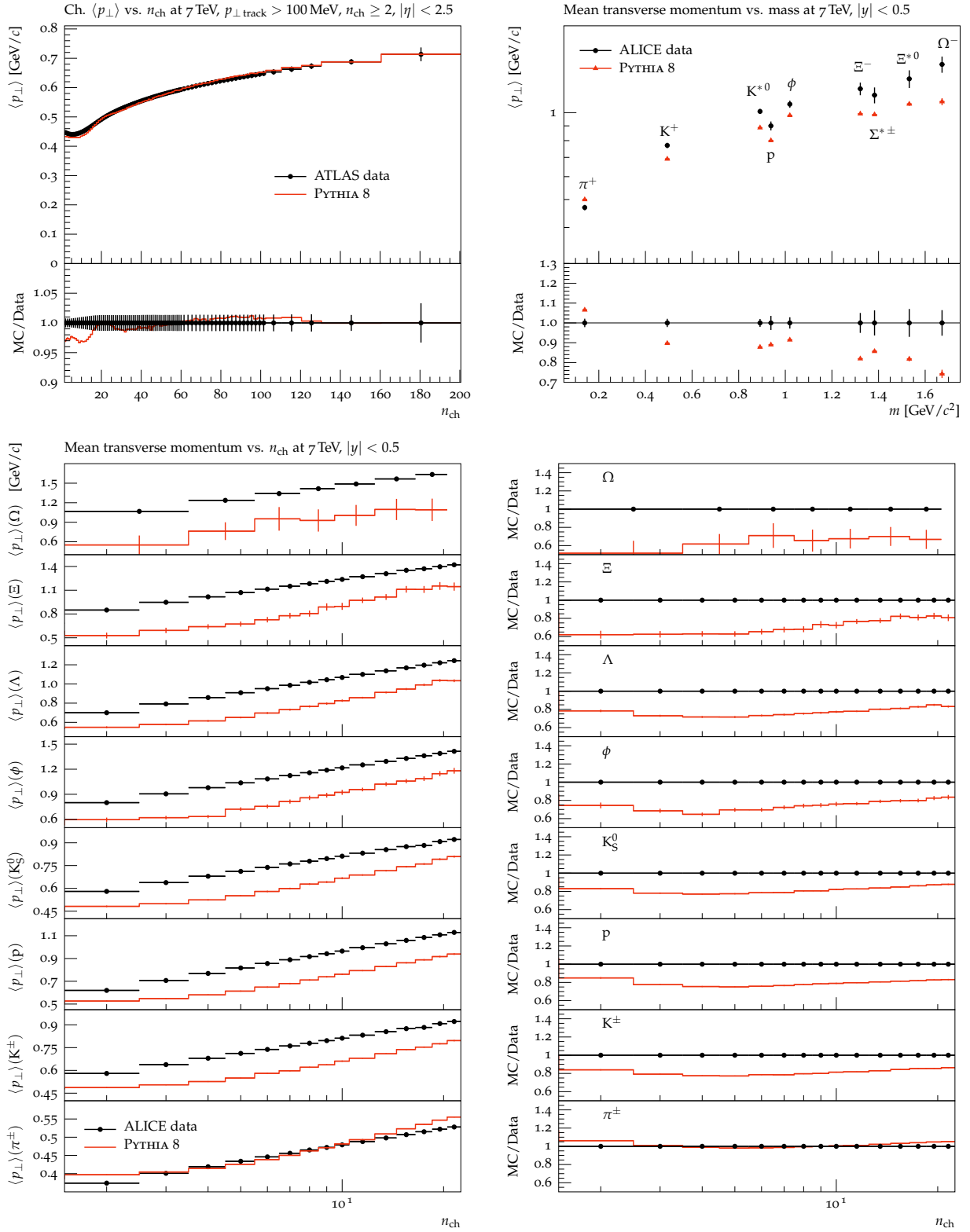


Figure 1: The mean transverse momentum as a function of the charged multiplicity (*top left* and the hadron mass (*top right* and *bottom*). Predictions of default PYTHIA compared to ALICE [45, 47] and ATLAS [46] data. The data in the bottom plots is taken to be an estimate of the logarithmic fits in [47] and therefore no error bars are included.

as follows [25]. The tail towards large n_{ch} is driven by events with more MPI activity, rather than e.g. by events with higher- p_{\perp} jets. If each MPI subcollision produces particles essentially independently the $\langle p_{\perp} \rangle(n_{\text{ch}})$ would be rather flat. CR implies that fewer and fewer extra particles are produced for each further MPI, as the possibilities to reduce the total string length by CR increase the more partons are already present. The amount of p_{\perp} from the MPIs thus increases faster than the n_{ch} , meaning more p_{\perp} per particle. (To this comes the normal hadronization p_{\perp} contribution, which raises the overall $\langle p_{\perp} \rangle$ level but does not contribute to the $\langle p_{\perp} \rangle(n_{\text{ch}})$ slope.) The exact nature of CR is not known, meaning that many models have been developed [25, 53, 54]. In most of them there is some overall CR strength parameter that can be adjusted to fit the $\langle p_{\perp} \rangle(n_{\text{ch}})$ slope.

- A natural reference for hadronization properties always is e^+e^- data. The principle of jet universality — or, in our case, string universality — is deeply rooted, so it is useful to check that no changes of fundamental string properties have too adverse an impact on e^+e^- . There is also a possibility of improvements in some places, like the inclusive $p_{\perp\text{in}}$ and $p_{\perp\text{out}}$ spectra; unfortunately these are not available for identified particles.

3 The New Models

In this section we outline the basic ideas and implementations that we have developed to offer new options to the traditional PYTHIA hadronization framework. As we later compare with data we will have reason to go into more detail and discuss some variations.

3.1 Variations of the normal string model

As described above, the standard tunneling framework suggests a Gaussian suppression of the production of heavier quarks and diquarks, with a further suppression based on the hadronic spin state, but no obvious room for an explicit dependence on the hadron mass. It also provides a common Gaussian p_{\perp} spectrum for all new $q\bar{q}$ pairs. We will study a few variations of this framework, mainly as a reference for the thermodynamical ansatz below.

Firstly, consider a Gaussian suppression associated with the masses of the produced hadrons rather than with the quarks. That is, let the relative production rate of different hadron species be given by a factor $\exp(-m_{\perp\text{had}}^2/2\sigma^2)$, which factorizes into a species-independent p_{\perp} spectrum and an $\exp(-m_{\text{had}}^2/2\sigma^2)$ mass suppression. The question is then whether this would give the appropriate suppression for the production of heavier particles.

Secondly, the universal p_{\perp} spectrum could be broken by assigning a larger width in string breakups of the $s\bar{s}$ and $qq\bar{q}\bar{q}$ kinds, relative to the baseline $u\bar{u}$ and $d\bar{d}$ ones. The issue to understand here is how dramatic differences are required to get a better description of the individual π , K and p p_{\perp} spectra.

Thirdly, assume that more MPIs leads to a closer packing of strings in the event, but that each string “flux tube” remains as a separate entity. The transverse region of the string shrinks and, essentially by Heisenberg’s uncertainty relations, this should correspond to a higher energy, i.e. a larger string tension κ . (Such a relation comes out naturally e.g. for bag models of confinement [55].) Overall the dense-packing effect on κ and related

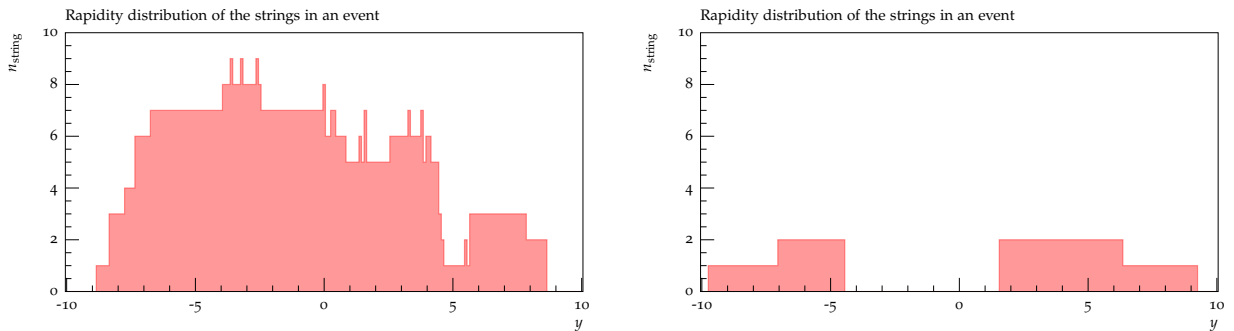


Figure 2: Rapidity distribution of the strings (added on top of each other) in a typical QCD event (*left*) and in a diffractive event (*right*).

parameters should scale as some power of n_{MPI} , i.e. the number of MPIs in the current event. Since n_{ch} and n_{MPI} are strongly correlated it is thus interesting to study how the particle composition and $\langle p_{\perp} \rangle$ depend on n_{ch} . For a more differential picture it should be preferable to estimate the number of strings in the neighbourhood of each new hadron being produced.

This is done by making a reasonable guess for the momentum of the hadron that is the next to be produced on the current string. Using an average hadron mass and p_{\perp} , defined in the frame of the parent string, and an average Γ value of $\langle \Gamma \rangle = (1+a)/b$, the momentum of the “average expected” hadron is calculated. Using this information, we determine the number of strings that cross the rapidity of the expected hadron. For this purpose the rapidity range that a string will populate is defined by the rapidity of the endpoint partons of each string piece,

$$y = \text{sgn}(p_z) \log \frac{E + |p_z|}{\sqrt{\max(m_{\perp}^2, m_{\min}^2)}} , \quad (6)$$

where m_{\min}^2 has the purpose to protect against strings with low- m_{\perp} endpoints from populating the full rapidity range. The rapidity-density measure is reasonable for low- p_{\perp} hadroproduction, but does not reflect the phase space inside a high- p_{\perp} jet, where close-packing of strings should be rare. Therefore the effective number of strings is calculated as

$$n_{\text{string}}^{\text{eff}} = 1 + \frac{n_{\text{string}} - 1}{1 + p_{\perp \text{had}}^2 / p_{\perp 0}^2} , \quad (7)$$

where $p_{\perp \text{had}}$ is the physical hadron p_{\perp} and $p_{\perp 0}$ is the MPI regularization parameter.

As two examples, the rapidity distribution of the strings in a typical QCD event and in a diffractive event are shown in fig. 2. Using eq. (7), the string tension in eq. (5) is modified to be

$$\kappa \rightarrow (n_{\text{string}}^{\text{eff}})^{2r} \kappa , \quad (8)$$

where the exponent r is left as a free parameter, that can be used to tune the model to data. Note that while junctions¹ contribute to the calculation of n_{string} by assuming one string stretched between the highest- and lowest-rapidity parton, their fragmentation does not make use of eq. (8). Junctions are rare in the models we study, so this is not a significant simplification.

¹A junction topology corresponds to an Y arrangement of strings, i.e. where three string pieces have to be joined up in a common vertex.

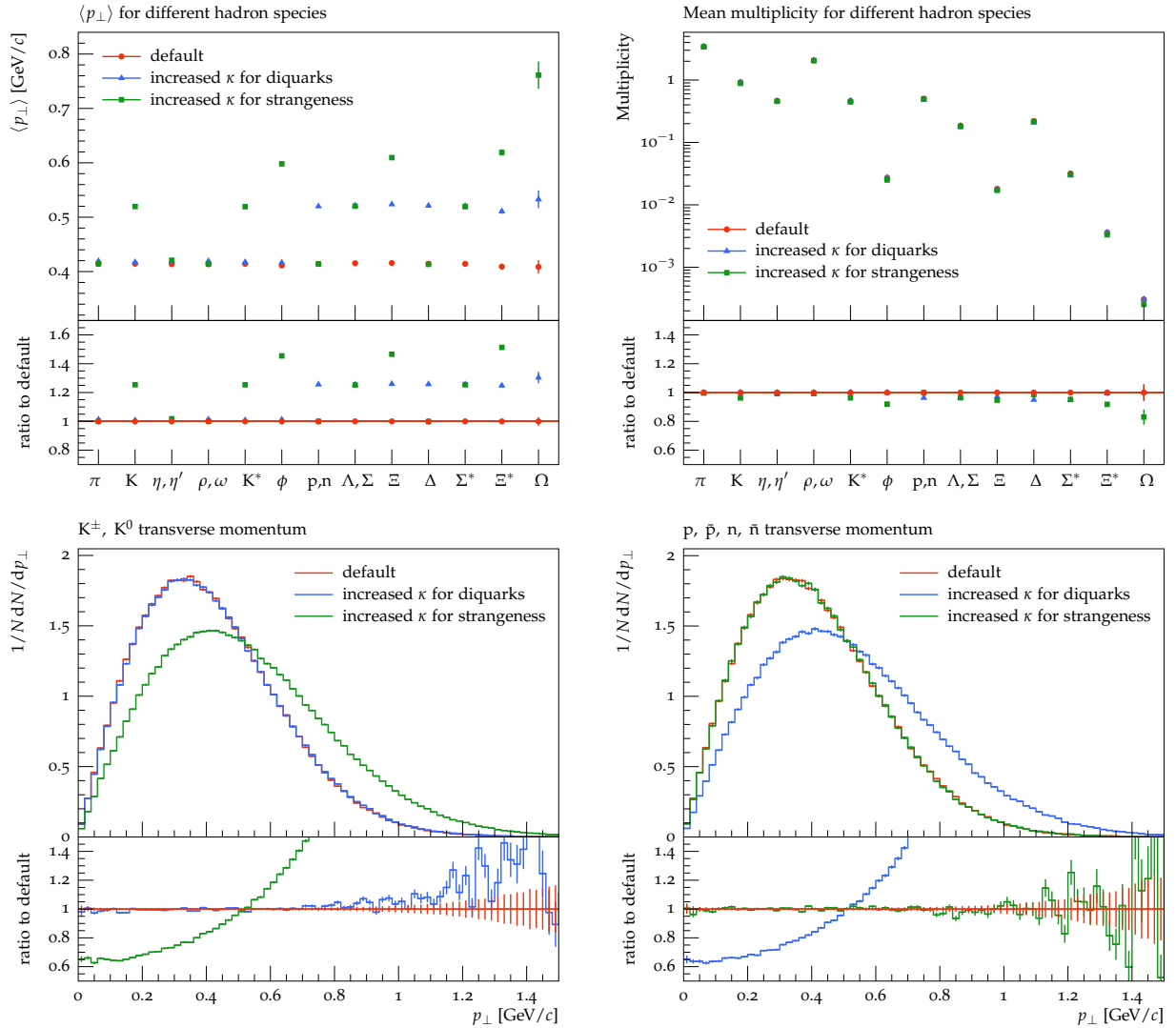


Figure 3: $\langle p_{\perp} \rangle$ (*top left*) and mean multiplicity (*top right*) for different hadron species and the K (*bottom left*) and p/n p_{\perp} (*bottom right*) spectra in the toy model. Predictions of the conventional string model without modifications are shown in red and with the string tension κ increased for diquarks in blue and strangeness in green.

3.1.1 One-string toy model

A very simple toy model is introduced to validate the modifications to the string tension in the conventional string model. A single string with energy m_Z is spanned along the z axis. The flavour of the endpoint quarks is chosen random from the set (u, d, s, c, b). The study includes only primary produced hadrons, i.e. no hadron decays, and also excludes the hadrons containing the endpoint quarks. (Such hadrons would have lower $\langle p_{\perp} \rangle$ since the endpoint quarks by definition have $p_{\perp} = 0$.)

The $\langle p_{\perp} \rangle$ and the mean multiplicity for different hadron species are shown in fig. 3. As expected, increasing the string tension either for s quarks or for diquarks leads to an increased $\langle p_{\perp} \rangle$ value for the hadrons concerned. Note that for $\eta + \eta'$ the $\langle p_{\perp} \rangle$ is only increased slightly due to the $u\bar{u} + d\bar{d}$ quark component being more frequently produced

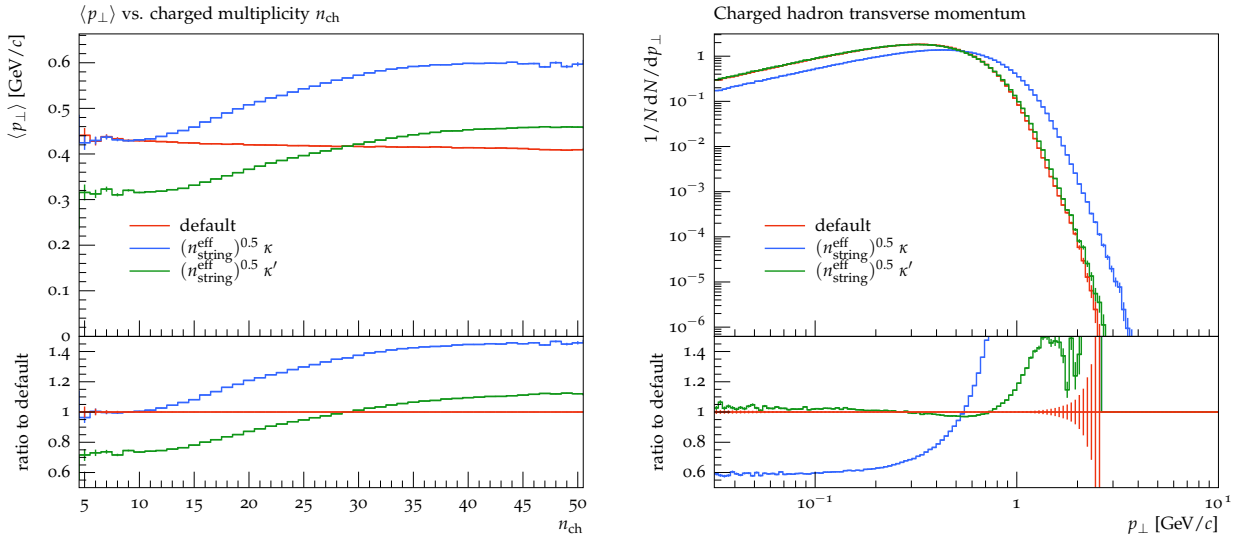


Figure 4: $\langle p_{\perp} \rangle$ as a function of the number of charged particles (*left*) and the p_{\perp} distribution (*right*) for the toy model with multiple strings along z axis. Predictions of the default model are shown in red and dependence of the string tension on the number of close strings in blue and green with two different string tensions $\kappa > \kappa'$.

compared to $s\bar{s}$. There is a slight reduction of the production probability for hadrons with s quarks or diquarks, shown in the top right plot in fig. 3, due to the increased string tension leading to fewer particles being produced in affected events. The bottom row of fig. 3 shows the K and p/n p_{\perp} spectra, shifted to larger values as the string tension for that hadron species is increased.

3.1.2 Multi-string toy model

To investigate the effect of the close-packing of strings, as in eq. (8), the above toy model is extended to include several strings along the z axis. The number of strings is picked randomly between two and eight and the string energies are chosen such that they sum up to 1 TeV. Fig. 4 shows $\langle p_{\perp} \rangle$ as a function of the number of charged particles and the p_{\perp} distribution and compares the modified model to default PYTHIA. Two different choices for the baseline value for the string tension are made in case of taking the close-packing of strings into account. In the first case the tension is denoted with κ and its value is adjusted such that $\langle p_{\perp} \rangle$ agrees with default PYTHIA for small values of n_{ch} . In the second case, where the string tension is denoted by κ' the value is adjusted to obtain the same $\langle p_{\perp} \rangle$, averaged over all hadrons and charged multiplicities. The latter case serves as a cross check when investigating the influence on the p_{\perp} spectrum of charged hadrons.

As expected the $\langle p_{\perp} \rangle$ increases with the charged multiplicity, eventually flattening out at large multiplicities. The left histogram in fig. 4 also nicely shows that the rise is independent of the baseline string tension value.

When fitting the string tension such that the same overall $\langle p_{\perp} \rangle$ is reached as in the default model, the charged hadron p_{\perp} spectrum exhibits only small changes; making the spectrum somewhat broader.

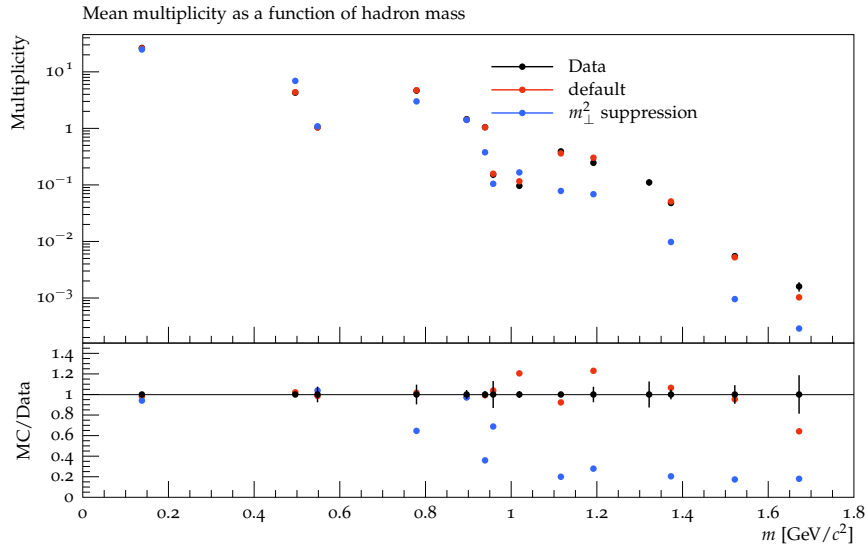


Figure 5: The conventional string model with its default options (red) and with the relative production rate of different hadron species given by a factor $\exp(-m_{\perp\text{had}}^2/2\sigma^2)$ (blue), compared to PDG data [56].

3.1.3 Gaussian $m_{\perp\text{had}}^2$ suppression

To test the applicability of the Gaussian transverse mass suppression, the quark p_{\perp} is generated according to $\exp(-p_{\perp q}^2/\sigma^2)$, see eq. (5), with the hadron flavour chosen based on $\exp(-m_{\perp\text{had}}^2/2\sigma^2)$. The additional factor of two arises from the hadron receiving p_{\perp} contributions from two quarks. As the comparison to data is of interest here, realistic $e^+e^- \rightarrow$ jets events with $s = m_Z^2$ are investigated. In fig. 5 the particle composition is shown as a function of mass. This clearly indicates that the suppression based on the transverse mass squared of the hadrons is suppressing heavier hadrons too much. We will therefore not consider this option further.

3.2 The thermodynamical string model

The most radical departure from standard Lund string principles that we explore in this article is to replace the Gaussian suppression factor in mass and p_{\perp} by an exponential one. To be more explicit, instead of a quark-level suppression governed by eq. (5) there will be a hadron-level suppression

$$\exp(-m_{\perp\text{had}}/T) \quad \text{with} \quad m_{\perp\text{had}} = \sqrt{m_{\text{had}}^2 + p_{\perp\text{had}}^2}. \quad (9)$$

The inspiration clearly comes from a thermodynamical point of view, which is why we choose to associate the dimensional parameter with a temperature T . This association should not be taken too literally, however; there are many differences relative to a purely thermal model. The main one is that we keep the longitudinal string fragmentation structure unchanged, which ensures local flavour conservation. Another is that e.g. the Hagedorn approach [18, 19] is based on the assumption of a steeply increasing density of excited states as a function of mass, whereas we only include a few of the lowest multiplets. (By

default only the ground states corresponding to no radial or orbital excitation, optionally also the lowest $L = 1$ meson multiplets.) This means that, although our T comes out to be a number of the order of the Hagedorn temperature, there is no exact correspondence between the two. Also, $T \sim \sqrt{\kappa/\pi} = \sigma$ from dimensional considerations, so our T could be viewed as a manifestation of the string energy per unit length, not directly linked to a temperature.

There is also an experimental historical background to the choice of an exponential shape, in that already fixed-target and ISR data showed that a distribution like $\exp(-Bp_\perp)$ offered a good fit to the inclusive $dn_{\text{ch}}/dp_\perp^2$ spectrum, with $B \approx 6 \text{ GeV}^{-1}$ [17, 57–59]. With data split by particle type, a lower B value is noted for kaons and protons than for pions, but with the modified form $\exp(-Bm_\perp)$ all the spectra can be described by almost the same $B \approx 6$ value.

As an aside, the preference for an exponential shape was and is not a show-stopper for the Gaussian approach in the normal string fragmentation. At larger p_\perp the spectrum is dominated by the fragmentation of (mini)jets, giving a larger rate than the nonperturbative hadronization one. And at smaller p_\perp the pattern of decays makes the spectrum more steep than the Gaussian one of the primary hadrons. So at the end of the day a Gaussian ansatz lands not that far away from an exponential spectrum, although differences remain. See further sec. 3.4.1, in particular fig. 9.

In more detail, our model is intended to give each new hadron in the string fragmentation a p_\perp according to an exponential distribution. We want to preserve the concept of local p_\perp conservation in each $q\bar{q}$ breakup vertex, so seek a distribution that convoluted with itself (in two transverse dimensions) gives an exponential,

$$f_{\text{had}}(p_{\perp\text{had}}) = \exp(-p_{\perp\text{had}}/T) = \int d^2p_{\perp 1} f_q(p_{\perp 1}) \int d^2p_{\perp 2} f_q(p_{\perp 2}) \delta(\mathbf{p}_{\perp\text{had}} - \mathbf{p}_{\perp 1} - \mathbf{p}_{\perp 2}) . \quad (10)$$

Unfortunately we have found no closed answer. Using Fourier transforms to turn the convolution into a product,

$$\begin{aligned} \tilde{f}_{\text{had}}(b_\perp) &= \frac{1}{2\pi} \int f_{\text{had}}(p_{\perp\text{had}}) \exp(-i\mathbf{b}_\perp \cdot \mathbf{p}_{\perp\text{had}}) d^2p_{\perp\text{had}} \\ &= 2\pi \tilde{f}_q^2(b_\perp) = \frac{1}{(1 + (b_\perp T)^2)^{3/2}} , \end{aligned} \quad (11)$$

and performing the azimuthal integration in the transformation back, leaves us with an expression

$$f_q(p_{\perp q}) \propto \int_0^\infty \frac{b J_0(b p_{\perp q}/T)}{(1 + b^2)^{3/4}} db , \quad (12)$$

where J_0 is a Bessel function and $b = b_\perp T$. In order to generate $p_{\perp q}$ values according to eq. (12), we divide the full $x_q \equiv p_{\perp q}/T$ spectrum into different ranges and use MATHEMATICA [60] to obtain numerical solutions and to fit functions of the form

$$\frac{\exp(a - c x_q)}{x_q^d} \quad (13)$$

to those solutions. We generate values of x_q according to $N \exp(-x_q)/x_q$, where N is adjusted such that eq. (13) is always overestimated, and use a simple accept-reject algorithm to obtain the x_q spectrum dictated by eq. (13).

Consider the fragmentation of a string, where the quark q of one breakup has a certain $\mathbf{p}_{\perp 1}$. The transverse momentum $\mathbf{p}_{\perp 2}$ of the (di)quark of the next breakup pair $q'\bar{q}'$ is constructed by picking its absolute value according to eq. (12) and a random azimuthal angle. The partner anti(di)quark must thus have $-\mathbf{p}_{\perp 2}$ due to local momentum conservation. The hadron transverse momentum is simply the sum of the p_{\perp} of the two contributing quarks, $\mathbf{p}_{\perp \text{had}} = \mathbf{p}_{\perp 1} - \mathbf{p}_{\perp 2}$. Having $p_{\perp \text{had}}$ at hand we decide on the flavour of the breakup pair $q'\bar{q}'$, and therefore also on the hadron species, as follows: calculate the transverse mass $m_{\perp \text{had}}$ of all hadrons whose flavour content includes the incoming quark q and determine the basic probability for each hadron as

$$P_{\text{had}} = \exp(-m_{\perp \text{had}}/T) . \quad (14)$$

Assuming the production of two hadrons with different masses m_1 and m_2 , then eq. (14) implies the same production rate for $p_{\perp} \gg m_1, m_2$, but more suppression of the heavier hadron at low p_{\perp} . Thus there is less production of heavier states, but they come with a larger $\langle p_{\perp} \rangle$.

As mentioned above, by default we only include the light-flavour (u, d, s) meson and baryon multiplets without radial or orbital excitation². However, if desired, more hadrons can be added to the procedure. Depending on the flavour content of the hadron, the probability in eq. (14) receives additional multiplicative factors:

- Due to spin-counting arguments vector mesons receive a factor of 3 and tensor mesons a factor of 5.
- For same-flavour mesons we include the diagonal meson mixing factors, similar to what has been done previously in the conventional Lund string model.
- Baryons receive a free overall normalization factor with respect to mesons, as well as an additional factor stemming from the $SU(6)$ symmetry factors, see [37]. The relative weight of spin 1/2 baryons with respect to those with spin 3/2 is 2 : 4, similar to the factors for mesons arising from the spin-counting arguments in point 1.
- For the special case of octet baryons with three different flavours, e.g. Λ and Σ^0 , their probability for different internal spin configurations is taken into account.
- An extra suppression factor for hadrons with strange (di)quarks is included to get more control over the relative hadron production and thus a better description of data.

All probabilities are then rescaled to sum up to unity and the hadron species and therefore the flavour of the next (di)quark pair is chosen accordingly. Note that we have not (yet) implemented popcorn baryon production, i.e. no mesons are produced in between a baryon and its antibaryon partner.

Similar to eq. (8) the temperature can be modified as

$$T \rightarrow (n_{\text{string}}^{\text{eff}})^r T , \quad (15)$$

with $n_{\text{string}}^{\text{eff}}$ given in eq. (7) to take into account the effect of close-packed strings.

²Heavy flavour hadrons are of course included to handle the endpoint quarks of the strings, where needed.

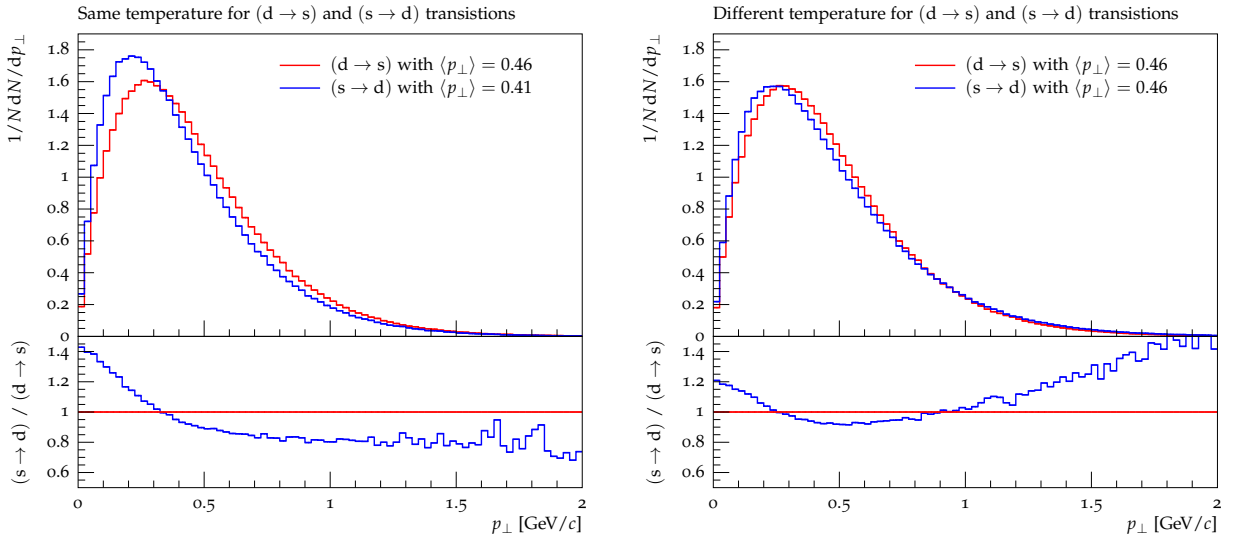


Figure 6: The inclusive hadron p_{\perp} spectrum for $(d \rightarrow s)$ and $(s \rightarrow d)$ transitions with the same temperature for both cases (*left*) and with the adjusted temperature (*right*).

3.2.1 Asymmetry in different flavour transitions

Consider a very simple model, where only string breaks with $d\bar{d}$ and $s\bar{s}$ quark pairs are allowed to produce only pseudoscalar mesons, and the mixing of diagonal mesons is ignored. Then it is rather easy to see that the p_{\perp} spectra and $\langle p_{\perp} \rangle$ of the hadrons produced in $(d \rightarrow s)$ transitions is not the same as for $(s \rightarrow d)$ transitions, due to the difference in competition. In the first instance $(d \rightarrow s)$ competes with $(d \rightarrow d)$, and since the former produces the heavier meson it also obtains the higher $\langle p_{\perp} \rangle$. In the latter instance $(s \rightarrow d)$ instead competes with $(s \rightarrow s)$ and so gives the lighter meson and lower $\langle p_{\perp} \rangle$. Assuming that fragmentation is performed from the quark end inwards, $K^0 = d\bar{s}$ would thus obtain a harder p_{\perp} spectrum than $\bar{K}^0 = \bar{d}s$, which should not be the case. A simple solution for obtaining the same $\langle p_{\perp} \rangle$ for both $(d \rightarrow s)$ and $(s \rightarrow d)$ transitions is to adjust the temperature in eq. (14) in case of initial s/\bar{s} quarks such that $d\bar{s}$ and $\bar{d}s$ hadrons are produced with the same $\langle p_{\perp} \rangle$ value, higher than $d\bar{d}$, and $s\bar{s}$ becomes even higher than that.

In fig. 6 we show the p_{\perp} spectra for both types of transitions with the same temperature and with the adjusted temperature. Note that though both transitions end up with the same $\langle p_{\perp} \rangle$ value, the shape of the p_{\perp} distribution still differs somewhat.

Unfortunately this is a price to pay for working with a recursive model, where flavour is conserved locally. A traditional thermal model based on eq. (14) would not conserve flavour or momentum, however, so is not an option here.

3.3 The hadronic rescattering model

A close-packing of fragmenting strings also implies a close-packing of the produced primary hadrons, i.e. a dense hadronic gas. This gives the possibility for hadrons to rescatter on the way out, in particular at the earliest times after hadronization. A detailed simulation of this mechanism would require a knowledge of where in space-time each hadron is produced. For a single string, say stretched along the z axis, it is straightforward to translate between the

(E, p_z) values of the primary hadrons and the (t, z) coordinates of the string breakups. For the more realistic case, when a string is stretched between several partons and the string motion is considerably more complicated [33], appropriate rules have not been worked out. To this should be added ambiguities in the transverse production coordinates, both as a consequence of the transverse distribution of the MPIs and of transverse fluctuations inside each string. The modelling of all of these aspects is an interesting task for the future. In addition, the cross section for the scattering of two hadrons against each other varies between hadron kinds, and depends on the relative energy of the two, adding a further layer of complexity.

Here we want to avoid such a detailed model, but still be able to explore whether hadronic rescattering effects could contribute to the resolution of some of the effects that we are attempting to explain. Collective flow — whether dictated by properties of the QGP or by hadronic rescattering — is well-established in heavy-ion collisions, see e.g. [61–63] and references therein. In particular, a common average radial velocity means that heavier particles have a higher $\langle p_\perp \rangle$ than lighter. Of course we do not expect as dramatic effects in pp, but they may still contribute to the same kind of $\pi/K/p$ p_\perp separation as in the thermodynamic scenario above, so it should be interesting to compare the two possibilities.

The simple modelling we have in mind is applied to the primary hadrons produced directly from the string fragmentation, before secondary decays are considered. Furthermore, for strings stretched along the z axis there is a strong correlation between the rapidity y of a particle and its space–time production vertex. Therefore, for a given hadron, the density of other hadrons at around the same rapidity is a reasonable (and longitudinally boost invariant) measure of how close-packed particle production is. If there is a contribution from particles coming from the same simple string it has presumably already been absorbed in the tuned fragmentation parameters, so we should disregard such pairs. Unlike e^+e^- events, however, it is common with topologies where a string consists of pieces stretched back and forth across the same rapidity range, and then the above argument does not apply. In practice, it is therefore more relevant to exclude rescattering only between close neighbours in the fragmentation chains.

One should further note that the rapidity density of hadrons refers to low- p_\perp particles. The hadronization of a scattered high- p_\perp parton mainly occurs at larger p_\perp scales, and these hadrons would be essentially unaffected.

The angular distribution of a rescattering, defined in the rest frame of the hadronic pair, should depend on the orbital angular momentum L . For simplicity, we restrict to s -wave isotropic scattering (i.e., $L = 0$) by requiring that the classical value of angular momentum $L = b|\mathbf{p}| < b|\mathbf{p}_{\max}| \sim 1$, where b is the impact parameter and \mathbf{p}_{\max} is the maximally allowed three-momentum of the hadrons in their rest frame, left as a (in principle) free parameter. We don't have access to b for each pair, but assume it is the same distribution for all combinations of hadron types. A common restriction on the three-momentum is thus introduced for all pairs, which is implemented as a cut on the invariant mass of the hadron pair,

$$m_{\text{inv}} < \sqrt{m_1^2 + |\mathbf{p}_{\max}|^2} + \sqrt{m_2^2 + |\mathbf{p}_{\max}|^2}, \quad (16)$$

with m_1 and m_2 being the masses of the hadrons and m_{inv} the physical invariant mass of the hadron pair. For all hadron pairs that are not excluded by eq. (16) we calculate the difference in rapidity, $\Delta y = |y_1 - y_2|$, and the rescattering probability. For hadrons that

are *not* produced in the same string the latter is

$$P_{\text{ds}}(\Delta y) = P_{\text{ds}}^{\text{max}} \left(1 - \frac{\Delta y}{\Delta y^{\text{max}}} \right), \quad (17)$$

where the maximum scattering probability $P_{\text{ds}}^{\text{max}}$ and the maximum rapidity difference Δy^{max} are left as free parameters. Eq. (17) simply means a probability of $P_{\text{ds}}^{\text{max}}$ for zero rapidity difference of the hadron pair, linearly decreasing to zero at a rapidity difference of Δy^{max} .

As an alternative to eq. (17), without the cut on the invariant mass in eq. (16), the probability can be chosen to be

$$P_{\text{ds}}(\Delta y, \Delta\varphi) = P_{\text{ds}}^{\text{max}} \left(1 - \frac{\sqrt{(\Delta y)^2 + (\Delta\varphi)^2}}{R^{\text{max}}} \right), \quad (18)$$

with $\Delta\varphi$ being the difference in azimuth of the hadron pair and R^{max} the maximally allowed value of the radius $R = \sqrt{(\Delta y)^2 + (\Delta\varphi)^2}$.

For hadron pairs that are produced in the *same* string we introduce the difference in hadron index (called rank e.g. in [64]), $\Delta_{ij} = |i - j|$, to denote how close two hadrons are, i.e. two neighbours have $\Delta_{ij} = \Delta_{i \ i+1} = 1$, next-to-neighbours $\Delta_{ij} = \Delta_{i \ i+2} = 2$ and so on. The scattering probability for same-string hadrons is

$$P_{\text{ss}}(\Delta y) = P_{\text{ds}}(\Delta y) \cdot \begin{cases} P_{\text{ss}}^{\text{max}} & \text{if } \Delta_{ij} > \Delta_{ij}^{\text{max}} \\ \frac{P_{\text{ss}}^{\text{max}}(\Delta_{ij} - \Delta_{ij}^{\text{min}}) + P_{\text{ss}}^{\text{min}}(\Delta_{ij}^{\text{max}} - \Delta_{ij})}{\Delta_{ij}^{\text{max}} - \Delta_{ij}^{\text{min}}} & \text{if } \Delta_{ij}^{\text{min}} \leq \Delta_{ij} \leq \Delta_{ij}^{\text{max}} \\ 0 & \text{if } \Delta_{ij} < \Delta_{ij}^{\text{min}}, \end{cases} \quad (19)$$

where $P_{\text{ss}}^{\text{min/max}}$ is the minimum/maximum probability associated with the nearest/furthest neighbour, characterized by $\Delta_{ij}^{\text{min/max}}$, with a linear behaviour of the probability in between; zero probability for hadrons closer than Δ_{ij}^{min} and maximum probability for those further apart than Δ_{ij}^{max} . All four are left as free parameters. In the case where eq. (18) is applied, $P_{\text{ds}}(\Delta y)$ in eq. (19) has to be replaced by $P_{\text{ds}}(\Delta y, \Delta\varphi)$.

3.3.1 Multi-string toy model

In order to test the hadron scattering a simple toy model is applied: five strings, each with energy m_Z , are constructed along the z axis with different quark flavours for the endpoint quarks, and the primary produced hadrons are studied. As in sec. 3.1 the hadrons containing the endpoint quarks are excluded. The following plots are obtained by making use of eqs. (16) and (17); using eq. (18) instead leads to very similar results, and thus the same conclusions.

Fig. 7 shows the $\langle p_{\perp} \rangle$ for different hadron species. For the Gaussian hadronic p_{\perp} distribution without hadron scattering all hadrons receive the same p_{\perp} spectrum. Including hadron scattering, the $\langle p_{\perp} \rangle$ decreases for pions, the lightest hadrons, by about 20% and increases for heavy hadrons by up more than 40%. The same effect is present for the thermodynamical model, although changes only reach around 10%, as the $\langle p_{\perp} \rangle$ is higher for heavier hadrons already without hadron rescattering.

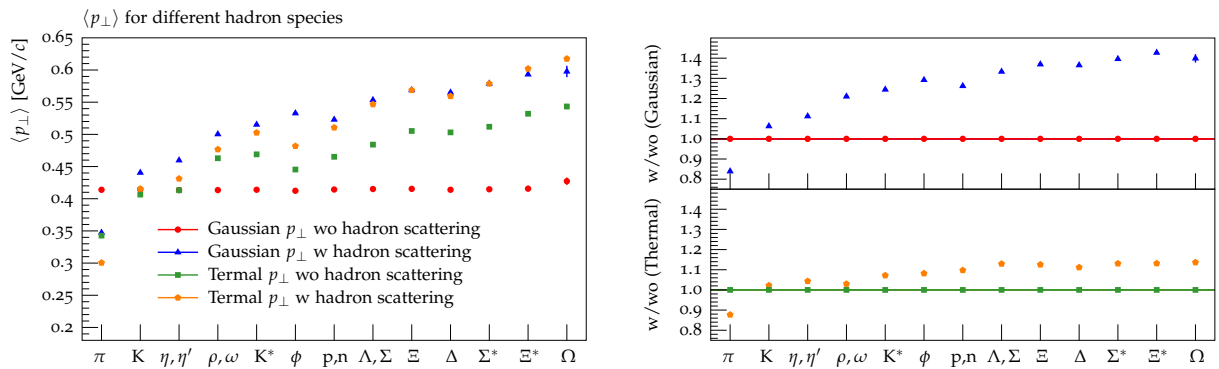


Figure 7: The $\langle p_{\perp} \rangle$ for different hadron species, with (blue triangles) and without (red dots) hadron scattering for the Gaussian p_{\perp} and with (orange pentagons) and without (green squares) hadron scattering for the thermal p_{\perp} in the toy model. The right plots show the ratio of the Gaussian model with to without hadron scattering in the upper panel and the same ratio for the thermodynamical model in the lower panel.

Fig. 8 shows the normalized p_{\perp} spectra for all hadrons and, to exemplify the difference between light and heavy hadrons, the spectrum for pions, kaons, and protons/neutrons. Comparing the inclusive p_{\perp} spectrum for the Gaussian p_{\perp} , we notice that the distribution gets broader with hadron scattering, i.e. we get more pions with small p_{\perp} and more heavy hadrons with higher p_{\perp} .

As the thermodynamical model without hadron rescattering comes with a different starting point, compared to the Gaussian model, the effect of the rescattering is not as large, and shifts the p_{\perp} spectrum towards larger values. The exception is for pions only, where there is a slight broadening also towards smaller p_{\perp} values.

3.4 Results

3.4.1 The Effect of Decays

Hadron decays, such as $\rho \rightarrow \pi\pi$ or $\eta \rightarrow \pi^+\pi^-\pi^0$, influence the p_{\perp} spectra of the final state hadrons. As the decays are mostly dictated by kinematics, they constitute a limiting factor on the possibilities of modifying for instance the pion p_{\perp} spectrum during the fragmentation process. Even though the primary hadrons follow a Gaussian or exponential p_{\perp} distribution, the spectra obtained after decays do not, and become more similar. In addition, in realistic events the effects of perturbative jet production leads to a p_{\perp} broadening and the emergence of a powerlike high- p_{\perp} tail.

To investigate this smearing of the p_{\perp} spectra, we consider realistic $e^+e^- \rightarrow$ jets events and inelastic pp collisions, where the previously discussed effects of string density and hadron rescattering are not taken into account. The normalized transverse momentum distributions of π^{\pm} and p, \bar{p} are shown in fig. 9 for the Gaussian and the thermodynamical model. Comparing the p_{\perp} spectra to the previous plots (note the different range of x and y axis) reveals how much the p_{\perp} distributions have moved to higher p_{\perp} values, as mentioned above. Four different ratio plots are included for each histogram to investigate different effects: the ratio of distributions after hadronic decays with respect to before, for both

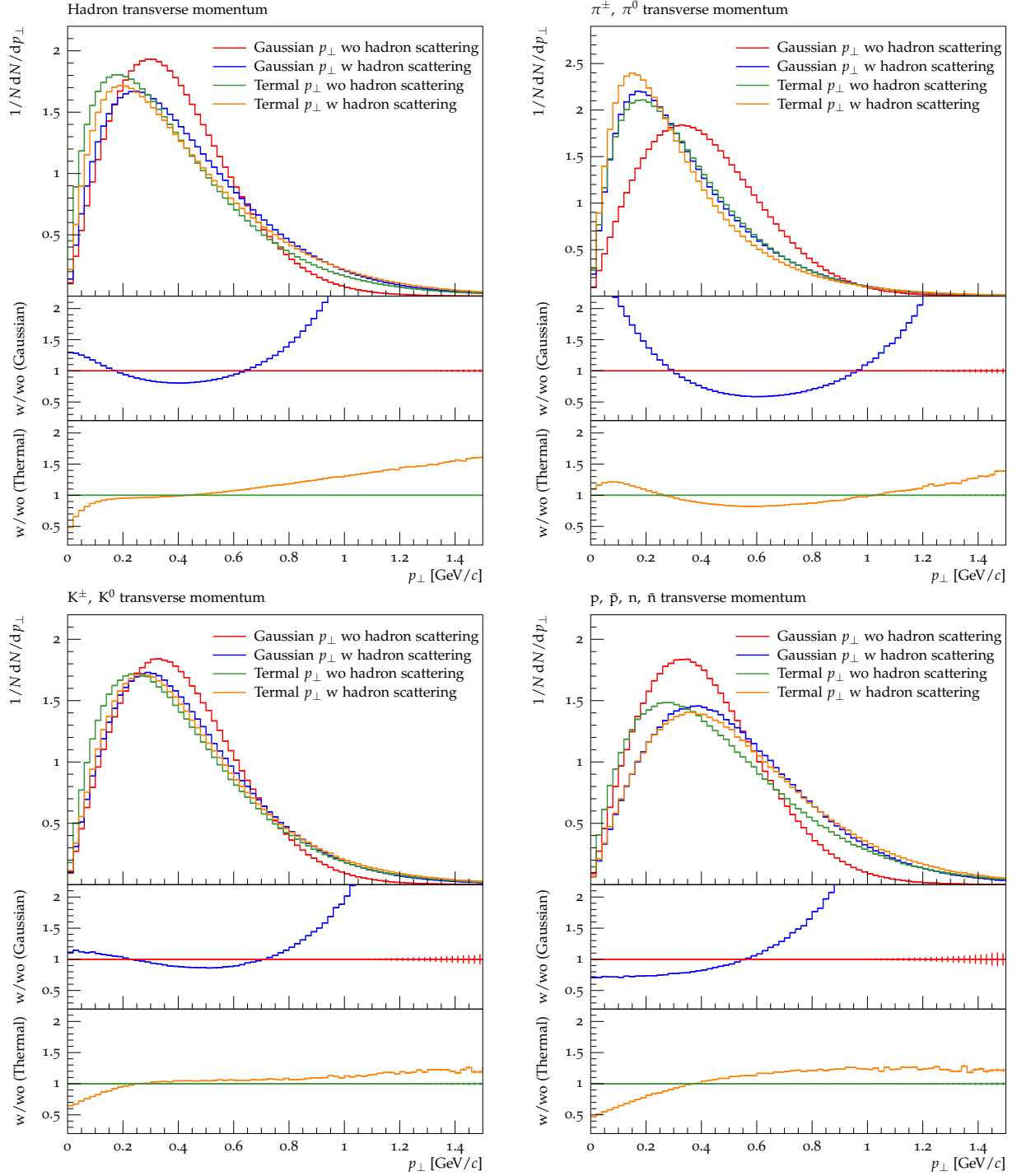


Figure 8: The p_{\perp} spectrum for all hadrons (*top left*), pions (*top right*), kaons (*bottom left*) and protons and neutrons (*bottom right*), with (blue) and without (red) hadron scattering for the Gaussian p_{\perp} and with (orange) and without (green) hadron scattering for the thermal p_{\perp} in the toy model.

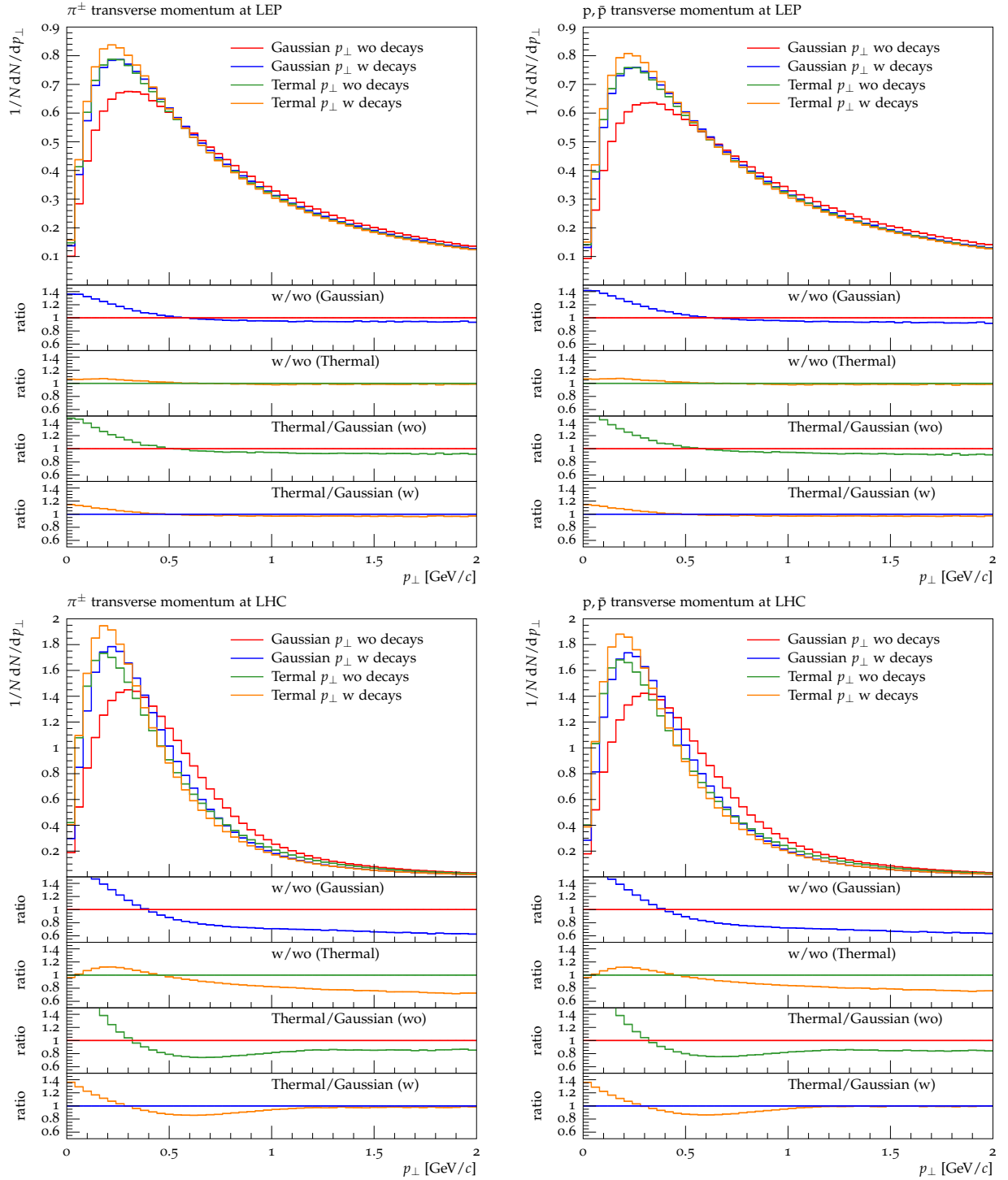


Figure 9: The p_{\perp} spectrum for charged pions (*left*) and protons (*right*), with (blue) and without (red) hadronic decays for the Gaussian and with (orange) and without (green) hadronic decays for the thermal model in $e^+e^- \rightarrow \text{jets}$ (*top*) and inelastic pp collisions (*bottom*).

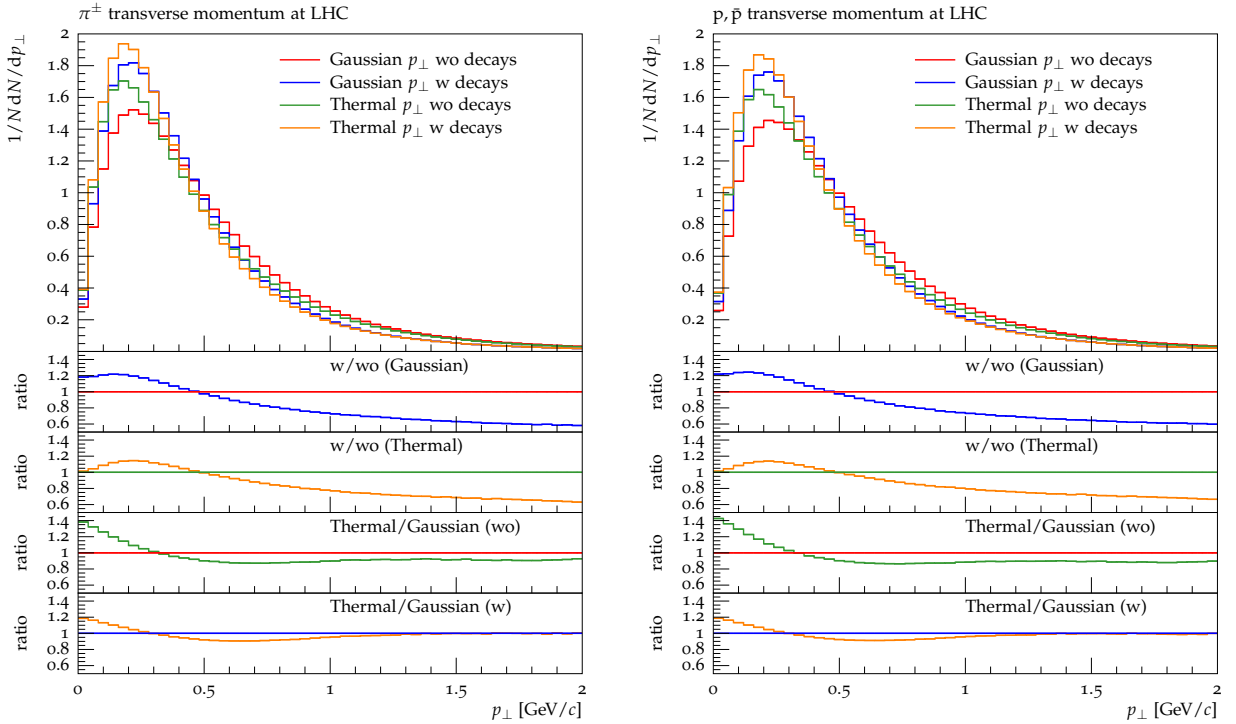


Figure 10: The p_{\perp} spectrum for charged pions (*left*) and protons (*right*), with (blue) and without (red) hadronic decays for the Gaussian and with (orange) and without (green) hadronic decays for the thermal model in inelastic pp collisions (*bottom*) with hadron rescattering.

models, and the ratio of the thermodynamical with respect to the Gaussian model, with and without decays.

Decays shift the p_{\perp} spectra towards smaller values, where the Gaussian model shows a larger change, compared to the thermodynamical model. For LEP the difference between the predictions with and without decays is limited to around 50% at most, while for LHC the changes are rather large, especially for small and large p_{\perp} values. Fig. 9 also nicely shows that the differences between the Gaussian and thermodynamical model become more than a factor of two smaller when hadronic decays are included, where, as before, the differences are more pronounced for LHC events. Unfortunately, this will limit the impact of the modifications previously discussed in this section.

Another question is how much of the hadron rescattering effect on the primary hadrons survives the decays. Note that some of the primary hadrons, such as the ρ meson, are so short-lived that some of their decay products could rescatter which would influence the p_{\perp} spectra further. A realistic interleaving of rescattering and decays would require a detailed space-time picture, however, which is for the future.

In fig. 10 the pion and proton p_{\perp} spectra are shown again for LHC events with hadron rescattering. The same ratio plots as before are included. The ratio of predictions with to without hadron decays shows a similar behaviour as the plots before, where no rescattering was included. The difference between the two models without hadronic decays becomes smaller when hadron rescattering is included. This is not surprising since the effect of the rescattering, that of shuffling some p_{\perp} from lighter to heavier hadrons, is smaller in the

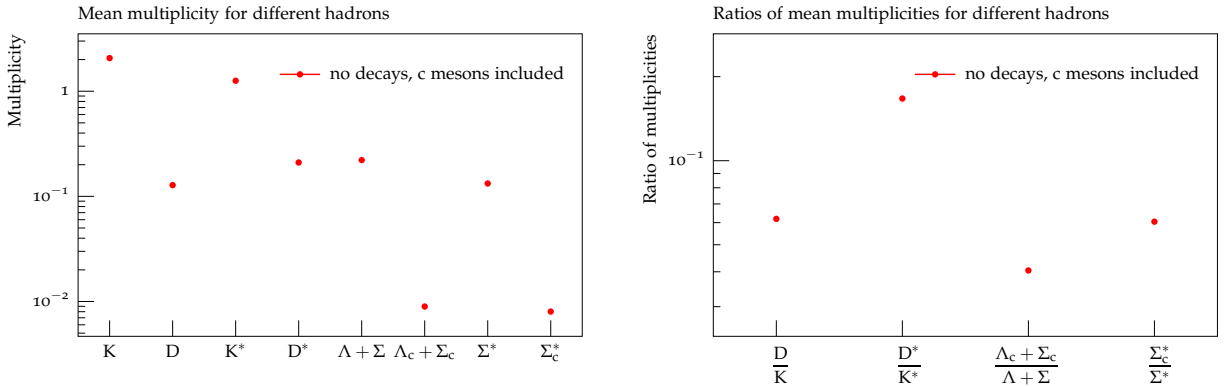


Figure 11: Mean multiplicities and their ratios for different groups of hadrons. Predictions of the thermodynamical model with c hadrons included in the fragmentation process and no hadronic decays.

thermodynamical model where more massive hadrons obtain more p_{\perp} already from the beginning. Including decays brings the predictions of the two models even closer together.

3.4.2 Adding more Hadrons

We now briefly investigate the effect of including additional hadrons in the flavour picking process in the thermodynamical model. $e^+e^- \rightarrow$ jets events at $\sqrt{s} = m_Z$ are analyzed with the effects of string density and hadron rescattering not being used. As discussed in sec. 3.2, by default only hadrons with u/d/s quarks and no radial or orbital excitation are included.

Firstly, consider including hadrons with charm quarks. To obtain a rough estimate of the suppression of c production in string breaks, compared to that of s quarks, the rates of D and K mesons and their ratios are analyzed; similar for vector mesons and baryons. The results in fig. 11 show that the c hadrons are suppressed by more than an order of magnitude compared to s hadrons, although a bit less when only vector mesons are considered. In absolute numbers the amount of extra charm production is non-negligible, and probably inconsistent with both LEP and LHC observed rates. Recall that an additional suppression factor for s quarks was introduced for the hadron rates in sec. 3.2; we would therefore expect that a similar, even stronger factor is needed when including c quarks in the thermodynamical model. Given this, neither charm nor bottom production is included in the nonperturbative hadronization in the rest of our studies.

Secondly, consider excited mesons by including all of the following meson multiplets, where by default only the first two are present,

- pseudoscalar multiplet with $L = 0, S = 0, J = 0$,
- vector multiplet with $L = 0, S = 1, J = 1$,
- pseudovector multiplet with $L = 1, S = 0, J = 1$,
- scalar multiplet with $L = 1, S = 1, J = 0$,
- pseudovector multiplet with $L = 1, S = 1, J = 1$,
- tensor multiplet with $L = 1, S = 1, J = 2$,

with J denoting the sum of the spin S and orbital angular momentum L in the nonrelativistic approximation. In fig. 12 the mean multiplicity of the different multiplets is shown, together with the p_{\perp} spectra of pions and protons. Note that including excited mesons

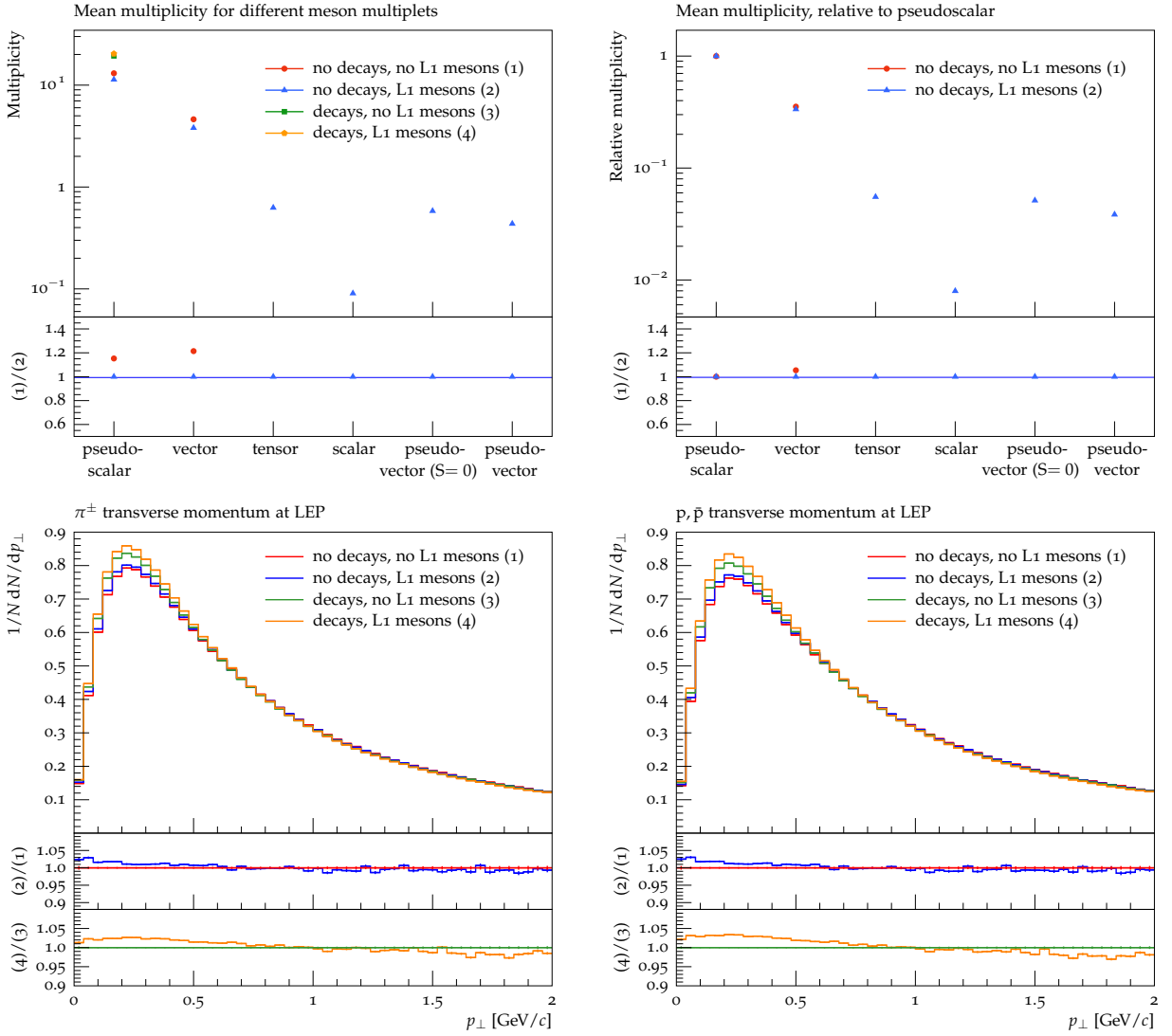


Figure 12: Mean multiplicities and ratios for the different meson multiplets (*top*). The p_{\perp} spectrum for charged pions (*bottom left*) and protons (*bottom right*). Predictions of the thermodynamical are shown, where decays are either switched on or off and $L = 1$ mesons are included or not.

leads to an increase of the total meson multiplicity after decays. All $L = 1$ multiplets are suppressed by more than an order of magnitude with respect to the pseudoscalar multiplet, with the scalar mesons being suppressed the most due the combination of them being the heaviest of the considered hadrons and their smaller spin-state weight $2J + 1$. The normalized p_{\perp} spectra exhibit slight shifts towards smaller values, as the now included heavier mesons decay to more lighter hadrons. The excited mesons combined constitute a fraction of roughly 10% of the total meson multiplicity. Given that in addition those mesons and their decay channels are not very well understood, we consider it reasonable to not include those in further studies. In default PYTHIA the suppression of light vector mesons with respect to pseudoscalar mesons is ~ 0.5 . The thermodynamical naturally comes with a fairly similar value of ~ 0.35 .

4 Comparisons with Data

We now proceed to compare the models with data. Note that, in a first step, there is no ambition to obtain a better overall description than the one achieved in several of the standard tunes that come with PYTHIA. It is rather to explore how the modelling of the new mechanisms impacts selected distributions, notably the ones discussed in sec. 2.2. That is, whether the mechanisms have the potential to improve the agreement with data in some crucial respects. Only in a second step is there some attempt to combine the various mechanisms, but still without the ambition of a full-fledged tune. In sec. 4.1 we present a comparison of the different effects we have discussed so far, while sec. 4.2 gives an overview of the results obtained by combining the effects into a more complete picture. Note that the new mechanisms will be available in the next public PYTHIA release.

4.1 Impact of the Different Effects

Based on a limited set of LHC observables, this section presents the impact of the new mechanisms outlined in sec. 3 on the description of data. The observables have been chosen to illustrate the effects of the change of the Gaussian width or temperature, respectively, depending on the close-packing of strings as in eqs. (8) and (15), of hadron rescattering, and of colour reconnection. The latter has been included as it serves a similar purpose and shows a somewhat comparable behaviour. The baseline prediction, which serves as the main comparison for both models, is obtained by switching off all of the aforementioned effects. For a clear picture of the influence of the individual mechanisms, only one of them is switched on at a time. Note that the prediction of the Gaussian model with colour reconnection, labelled “Gaussian p_{\perp} ColRec” in the plots, corresponds to default PYTHIA 8. Recall that the results presented in this subsection are not obtained with parameter settings that optimize the data description but rather illustrate their general effect. The average transverse momentum $\langle p_{\perp} \rangle$ as a function of the hadron mass and the charged multiplicity are shown in fig. 13, together with the charged particle p_{\perp} spectrum.

For both models, the description of $\langle p_{\perp} \rangle$ as a function of mass improves for each of the different mechanisms, compared to the baseline prediction, as heavier hadrons obtain larger $\langle p_{\perp} \rangle$ values. The thermodynamical model provides a somewhat better description of this observable, compared to the Gaussian model, which comes naturally due to the exponential hadronic transverse-mass suppression.

The baseline prediction for $\langle p_{\perp} \rangle(n_{\text{ch}})$ plateaus at small multiplicities, therefore underestimating $\langle p_{\perp} \rangle$ for values $n_{\text{ch}} \gtrsim 25$. All of the effects investigated in this study have a somewhat similar effect, in the sense that they are able to push up the prediction, compared to the baseline settings. While including the $n_{\text{string}}^{\text{eff}}$ -dependence significantly improves the description, it is still slightly worse than the prediction with colour reconnection. The hadron rescattering provides a fairly good description of $\langle p_{\perp} \rangle$ for small n_{ch} values, but clearly overshoots the distribution at high multiplicities.

Similar to the previous observable, the $n_{\text{string}}^{\text{eff}}$ -dependence and colour reconnection improve the description of the inclusive p_{\perp} spectrum. The Gaussian model without additional effects switched on produces a bump at $p_{\perp} \sim 0.5$ GeV/ c and a broad dip at $p_{\perp} \sim 2.5$ GeV/ c . While colour reconnection removes the dip almost completely, the bump is still clearly visible. The $n_{\text{string}}^{\text{eff}}$ -dependence somewhat reduces both bump and dip, but at the cost of

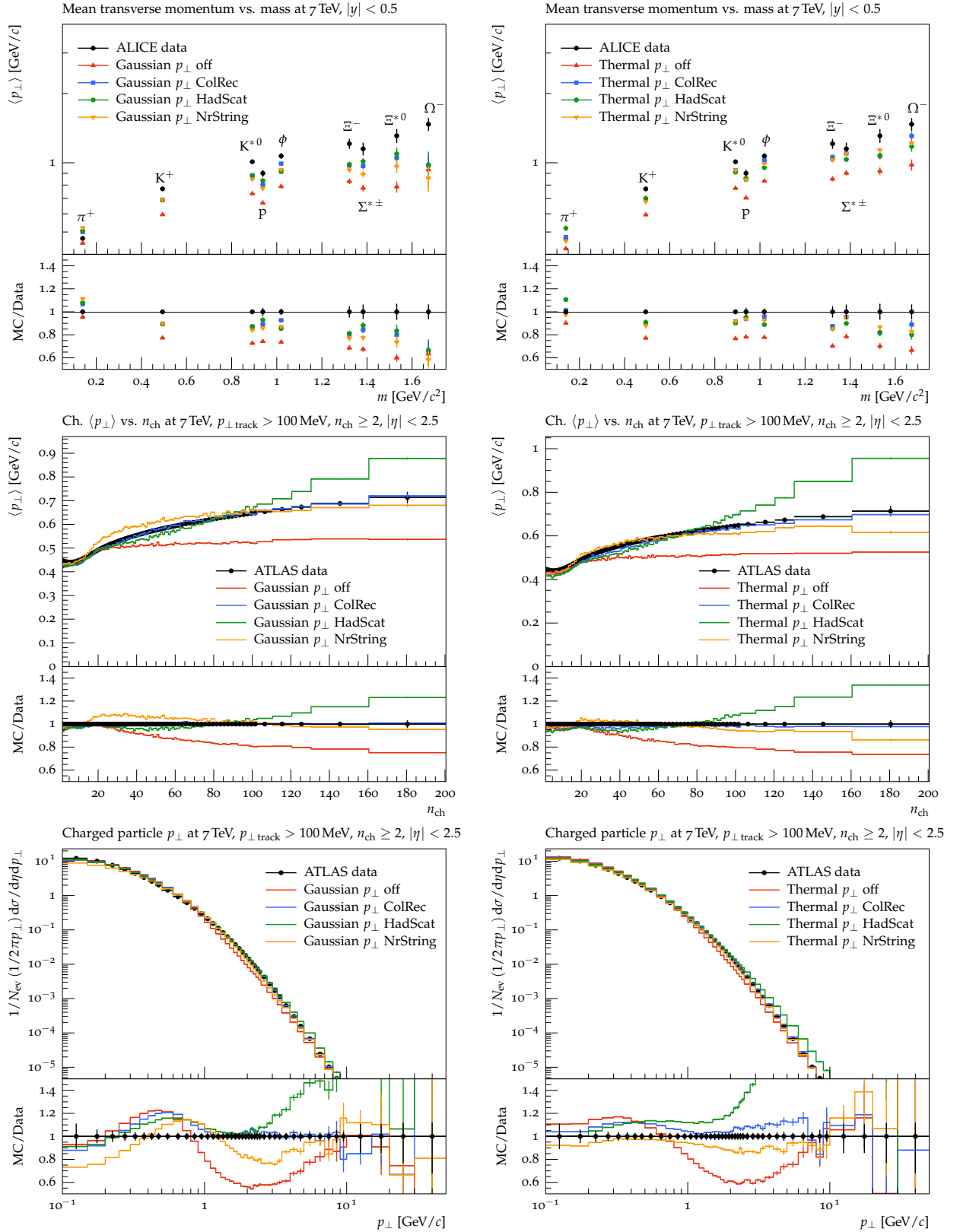


Figure 13: Comparisons to ALICE [45] and ATLAS [46] data: $\langle p_{\perp} \rangle$ as a function of the hadron mass (*top*), charged multiplicity (*middle*), and the charged particle p_{\perp} (*bottom*). Predictions with the Gaussian (thermodynamical) model are shown in the *left* (*right*) plots. ColRec/HadScat/NrString means that only colour reconnection/hadron rescattering/ $n_{\text{string}}^{\text{eff}}$ -dependence is switched on, otherwise everything is switched off.

introducing another dip towards very small p_{\perp} values. The baseline prediction of the thermodynamical model, compared with the Gaussian one, has the same dip at $p_{\perp} \sim 2.5 \text{ GeV}/c$, while the bump is much less visible. Both colour reconnection and the $n_{\text{string}}^{\text{eff}}$ -dependence reduce the dip quite substantially and provide a very good description of the data. The hadron rescattering, while improving the description in the low- p_{\perp} region, overshoots the distribution quite significantly for high- p_{\perp} values in both models.

4.2 Results

Using the information of the last subsection we adjust the parameters associated with the new mechanisms to obtain a good data description, with the main focus lying on p_{\perp} spectra of pions, kaons, and protons. We begin with LHC data, as this is the motivation for the thermodynamical model, the $n_{\text{string}}^{\text{eff}}$ -dependence, and the hadron rescattering, and continue with a cross-check of some LEP and SLC observables.

4.2.1 LHC

The new parameters are adjusted such that an improvement of the p_{\perp} spectra of π^{\pm} , K^{\pm} and p, \bar{p} , measured with ALICE [50], is achieved, while still giving a reasonable description of the charged particle p_{\perp} distribution and $\langle p_{\perp} \rangle$ as a function of the multiplicity, both measured with ATLAS [46]. The corresponding settings and values can be found in app. A. The LHC data set presented here includes the aforementioned p_{\perp} spectra and the Λ to K_S^0 ratio shown in fig. 14, $\langle p_{\perp} \rangle$ as a function of the hadron mass and the charged multiplicity, both inclusive and for different hadrons, shown in fig. 15, and the ratio of yields with respect to $(\pi^+ + \pi^-)$ as a function of the charged multiplicity for different hadrons, shown in fig. 16. The predictions of default PYTHIA are compared to the Gaussian and thermodynamical model with the modifications outlined in sec. 3.

Default PYTHIA describes the ATLAS charged particle p_{\perp} distribution very well for values of $p_{\perp} > 1 \text{ GeV}/c$, but shows a bump at around $0.5 \text{ GeV}/c$. The Gaussian model with modifications gives a similar shape and reduces the bump somewhat, while undershooting the distribution large p_{\perp} by a few %. The thermodynamical model improves the description quite substantially, especially for low- p_{\perp} values, where the aforementioned bump is almost gone. The predictions for the CMS charged hadron p_{\perp} spectrum behave mostly similar, with the same bump visible for default PYTHIA and the Gaussian model.

For default PYTHIA, pions obtain a too hard p_{\perp} spectrum. The modifications to the Gaussian model improve the distribution slightly, but there is still no good overall description. With the thermodynamical model the spectrum improves for low- p_{\perp} values quite a bit; however, it is still a bit too high in the large- p_{\perp} region. The K^{\pm} p_{\perp} spectrum shows the opposite behaviour: too many soft and too few hard kaons. The Gaussian model with modifications improves the description in the soft region somewhat, compared to default PYTHIA. Both the Gaussian and thermodynamical model change the shape of the spectrum slightly, but do not provide a better overall description of the K^{\pm} p_{\perp} . The prediction of both models for the p, \bar{p} p_{\perp} spectrum are better compared to default PYTHIA, where especially the thermodynamical model improves the low- p_{\perp} region quite substantially.

While the prediction of the thermodynamical model for the Λ/K_S^0 ratio is somewhat flatter with respect to the data, especially in the low- p_{\perp} region, the normalization is off by

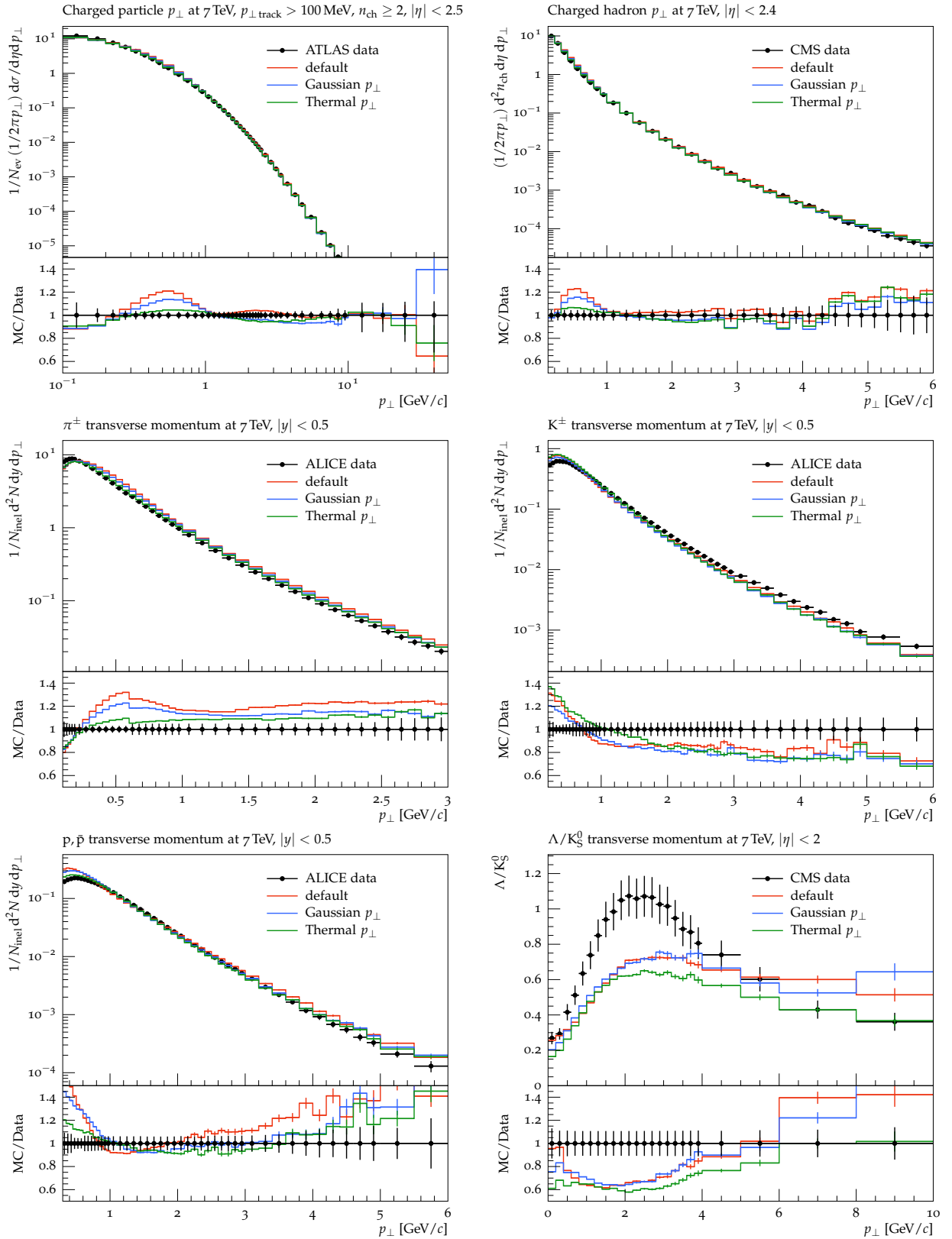


Figure 14: Inclusive (*top*), π^{\pm} , K^{\pm} , and p, \bar{p} (*middle and bottom left*) p_{\perp} spectra and the Λ to K_S^0 ratio (*bottom right*). Predictions of default PYTHIA, the Gaussian and thermodynamical model with modifications, compared to ATLAS [46], CMS [51, 65] and ALICE data [50].

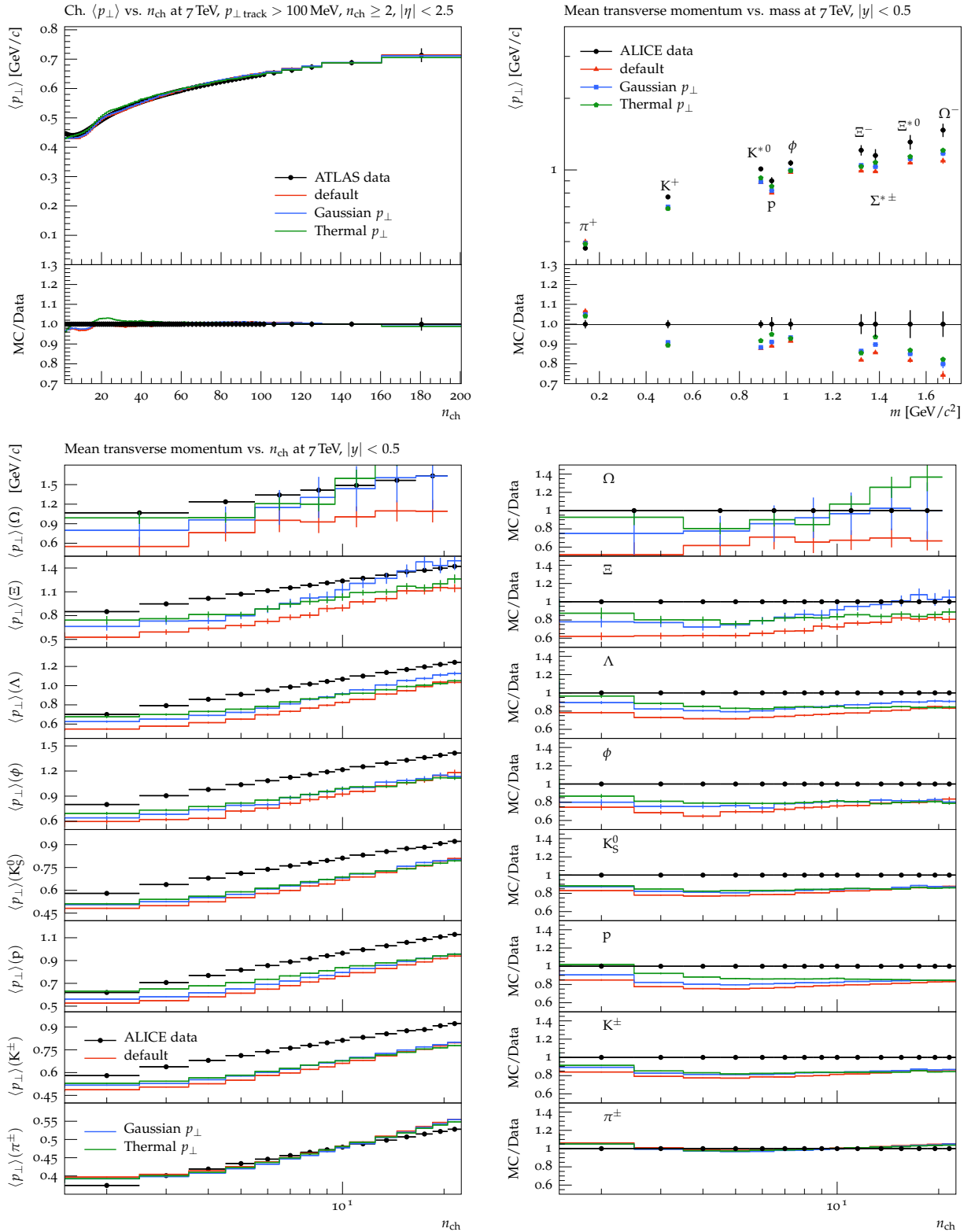


Figure 15: The mean transverse momentum as a function of the charged multiplicity (*top left*) and the hadron mass (*top right*) and (*bottom*). Predictions of default PYTHIA, the Gaussian and thermodynamical model with modifications, compared to ALICE [45,47] and ATLAS [46] data. The data in the bottom plots is taken to be an estimate of the logarithmic fits in [47] and therefore no error bars are included.

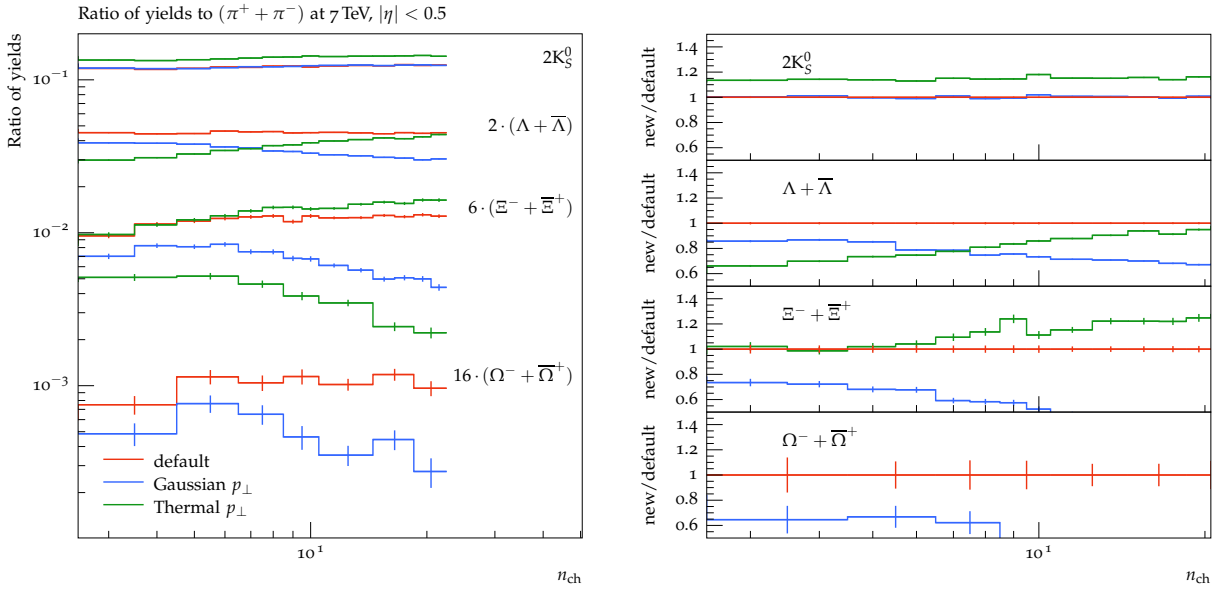


Figure 16: Ratio of yields with respect to $(\pi^+ + \pi^-)$ as a function of the charged multiplicity. Predictions of default PYTHIA, the Gaussian and thermodynamical model with modifications. The ALICE measurement can be found in [10].

almost a factor of two due to the combination of producing slightly too many K_S^0 and not enough Λ . The observable could be improved by adjusting the overall normalization factor of baryons with respect to mesons. The value of this parameter has been fixed using the proton p_\perp spectrum, however.

All models give very similar predictions for $\langle p_\perp \rangle$ as a function of n_{ch} , with an extremely good description of the region $n_{\text{ch}} > 20$, but too low $\langle p_\perp \rangle$ for smaller multiplicities. It is quite obvious that the description of this observable could be further improved by choosing a larger value for the width or temperature respectively and simultaneously lowering the $n_{\text{string}}^{\text{eff}}$ -dependence and hadron rescattering. This would however come hand in hand with worse descriptions of other observables. For default PYTHIA, pions obtain a too large $\langle p_\perp \rangle$ and heavier hadrons a too small one. While the thermodynamical model improves the predictions for $\langle p_\perp \rangle(m)$, there is still no full agreement with data. The Gaussian model lies in between default PYTHIA and the thermodynamical model. We observe a similar behaviour for the $\langle p_\perp \rangle(n_{\text{ch}})$ distribution for individual hadrons. The pion $\langle p_\perp \rangle$ is described fairly well with a slope that is slightly too steep. The main difference of the other hadrons with respect to pions is that they obtain a too small $\langle p_\perp \rangle$ over the whole n_{ch} range. As for pions, the slopes tend to be too steep.

ALICE [10] found that the production of strange and multi-strange hadrons is enhanced with increasing multiplicity. While default PYTHIA is not able to reproduce such a behaviour, fig. 16 shows that the thermodynamical model achieves an increase of strangeness with charged multiplicity for K_S^0 , Λ , and Ξ , but not for Ω . Except for the latter, we therefore expect the thermodynamical model to give an improved description of the data presented in [10]. The Gaussian model with modifications shows the opposite effect, a decrease with growing multiplicity. These findings can be explained as follows: In default PYTHIA all (primary) hadrons are produced with a probability that is independent of the multiplic-

ity or number of strings. In the thermodynamical model heavier hadrons are produced preferably at large p_{\perp} values. Including the $n_{\text{string}}^{\text{eff}}$ -dependence leads to potentially higher temperatures for events with large n_{ch} , where heavy hadrons have a higher probability to be produced, compared to low- n_{ch} events. With the modifications to the Gaussian model, all hadrons obtain more p_{\perp} in events with large values of $n_{\text{string}}^{\text{eff}}$. Due to phase-space constraints heavier hadrons might be rejected more often compared to lower-mass hadrons, leading to the decrease with growing multiplicity. This might also be the reason for the slight drop towards large n_{ch} for Ω in the thermodynamical model, as it eventually dominates over the effect of the $n_{\text{string}}^{\text{eff}}$ -dependence.

4.2.2 LEP and SLC

While the main motivation for introducing the exponential p_{\perp} distribution is arising from LHC data, the valid question of whether the same model is able to describe e^+e^- observables as well remains. The effect of the close-packing of strings and hadron rescattering are not included for e^+e^- data as we do not expect them to represent relevant physics here. Furthermore the string dependence relies on rapidity differences and an event axis aligned with the beam, which is not present in e^+e^- collisions.

The parameters of the thermodynamical model are adjusted using the charged particle momentum spectrum as well as the scaled momenta of π^{\pm} , K^{\pm} and p, \bar{p} , measured with SLD [66], while the ALEPH event shapes [67] served as cross checks. For the Gaussian model the width and prefactors for strange and diquarks have been adjusted such that the mean charged multiplicities agrees with the value obtained with the thermodynamical model. The corresponding settings and values can be found in app. A. The e^+e^- data set presented here includes the aforementioned momenta and mean multiplicities for different hadrons shown in fig. 17, as well as the charged multiplicity distribution, scaled momentum and the inclusive $p_{\perp\text{in}}$ and $p_{\perp\text{out}}$ spectra shown in fig. 18. The predictions of default PYTHIA are compared to the Gaussian and thermodynamical model as described above.

The only difference between default PYTHIA and the prediction labelled as ‘‘Gaussian p_{\perp} ’’ is an adjusted value for the Gaussian width and its prefactors for s and diquarks, i.e. there is no change of the flavour selection parameters. Therefore, the values of the mean multiplicities remain, leading to overlapping data and Monte Carlo histogram points in fig. 17, which are thus not fully visible. With the thermodynamical model we obtain a fairly good description of most hadrons, with the notable exceptions of producing too many heavy baryons. Note however, that the Gaussian model comes with around 20 parameter for selecting the flavour of new hadrons, whereas the thermodynamical model makes use of only three parameters: the temperature, the overall normalization factor of baryons with respect to mesons, and the additional suppression factor for hadrons with strange quarks, see sec. 3.2. Hence, the result is fairly acceptable.

The predictions of the two models for the charged particle momentum agree very well with data in the soft region; there is only some small deviation for medium and large momenta, where especially the thermodynamical model predicts somewhat too many particles in the hard region. The same effect is even clearer visible in the scaled momentum spectrum of pions. For kaons and protons we observe the opposite effect: the new model predicts too few hadrons with large momenta.

Similar to the charged particle momentum, the predictions of both models for the loga-

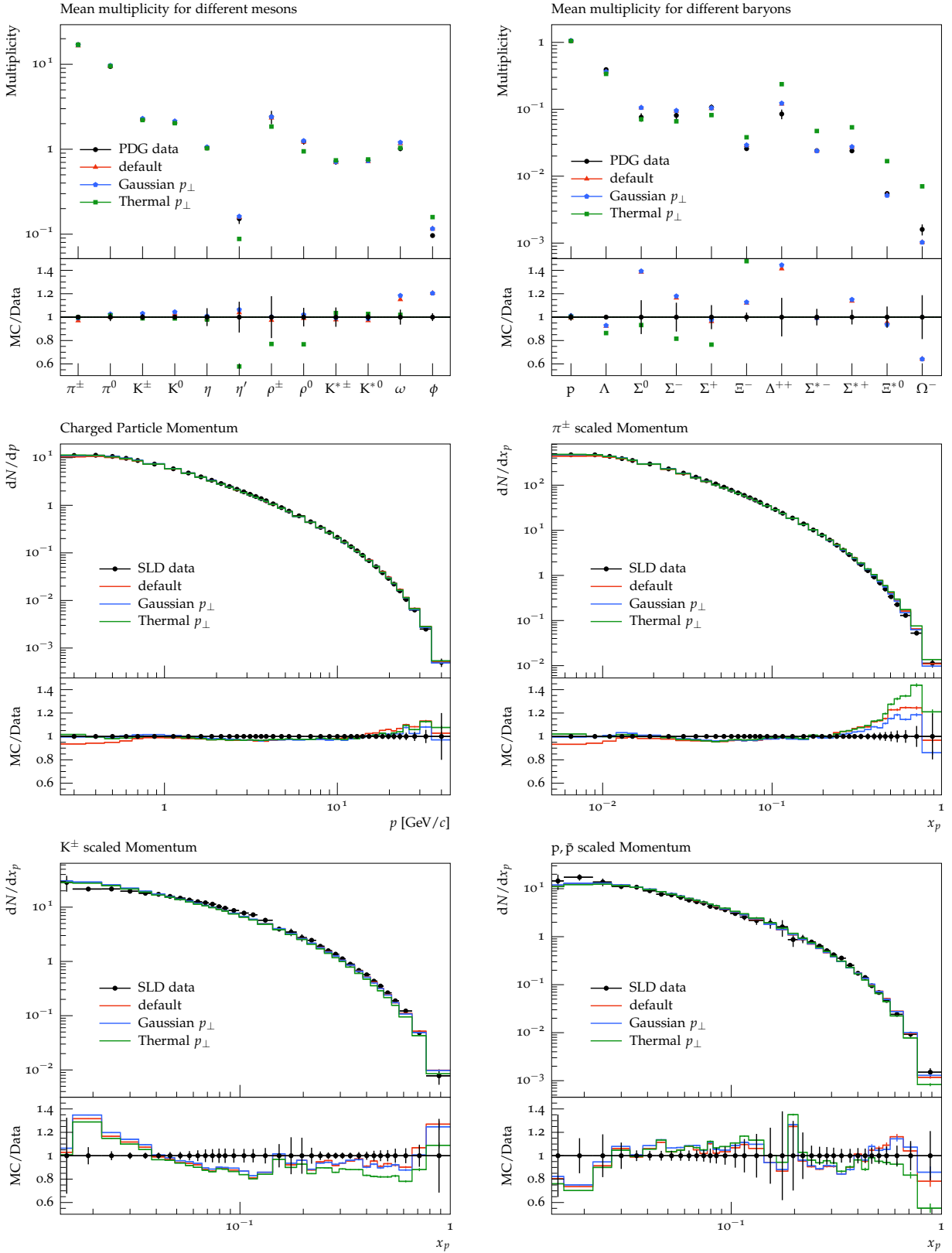


Figure 17: Mean hadron multiplicities (*top*), charged particle momentum (*middle left*), and scaled momenta $x_p = 2|\mathbf{p}|/E_{\text{cm}}$ of π^\pm (*middle right*), K^\pm (*bottom left*) and p, \bar{p} (*bottom right*). Predictions of default PYTHIA 8, the Gaussian and thermodynamical model compared to PDG [56] and SLD data [66].

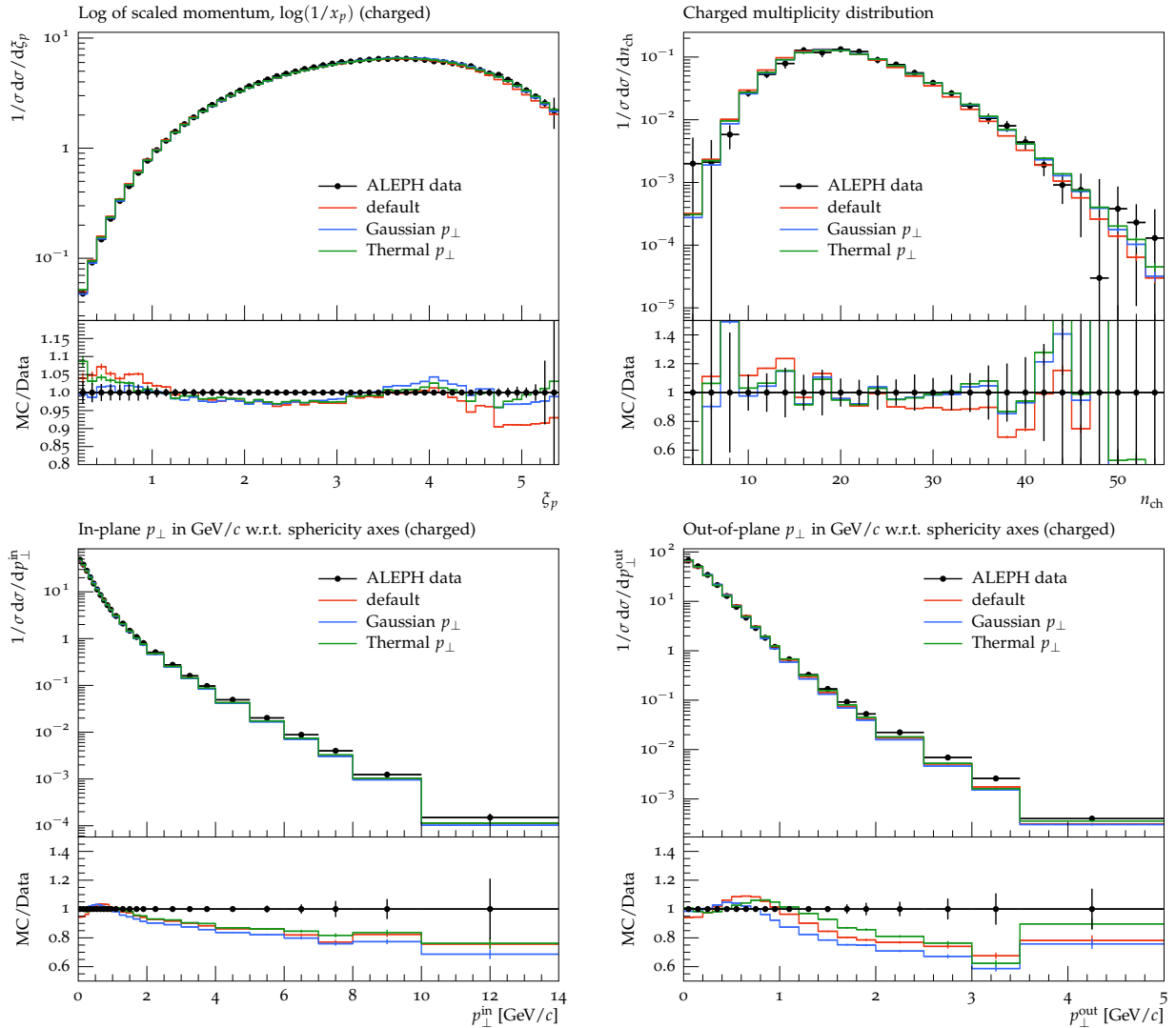


Figure 18: Charged multiplicity distribution (*top left*), scaled momentum (*top right*) and the inclusive $p_{\perp\text{in}}$ and $p_{\perp\text{out}}$ spectra (*bottom*). Predictions of default PYTHIA 8, the Gaussian and thermodynamical model compared to ALEPH data [67].

rihm of the scaled momentum agrees well with data, with some small deviation for medium and large values. However, the sudden drop in the ratio of the Monte Carlo prediction to data at around $\xi_p = 4.7$ remains almost unchanged. The description of the charged multiplicity distribution improves slightly, compared to default PYTHIA, towards having less events with small multiplicities and more events with larger ones. This is due to having an increased mean charged multiplicity. While both models describe the low- p_{\perp} region of the inclusive $p_{\perp\text{in}}$ and $p_{\perp\text{out}}$ spectra very well, they underestimate the amount of events with larger p_{\perp} values. The thermodynamical model provides a better description of especially the $p_{\perp\text{out}}$ spectrum, compared to the Gaussian model.

To summarize we note that the thermodynamical model is able to provide predictions for event shapes and momentum spectra in e^+e^- events that are of a similar quality as those by the Gaussian model. Nevertheless, the hadron decomposition is not described well, a price to pay for reducing the amount of flavour selection parameters.

5 Summary and Outlook

The understanding of soft hadronic physics is changing under the onslaught of LHC pp data. Of course, there has never been an approach that could describe all aspects of pp physics perfectly, but before LHC it was often assumed that all the basic concepts were in place, and that what remained was successive refinements. Now we see that there is still much left to learn. There have already been several surprises, and further data analyses may well produce more.

In view of this we have revisited some of the basic soft-physics assumptions of the PYTHIA event generator, which has been quite successful in predicting and describing many aspects of the data, but now starts to show cracks. New approaches have here been studied for some areas, to understand how much room for improvements there would be, without any claim that either of them would necessarily be the one and only right way to go.

A central pillar of PYTHIA has been the Lund string fragmentation model, where a tunneling mechanism for string breakups leads to a universal Gaussian p_{\perp} spectrum. In this work a thermodynamical model is implemented as an alternative, where p_{\perp} instead follows an exponential distribution. For an already selected p_{\perp} , the hadron flavour is picked based on an exponential m_{\perp} weight, with additional factors due to spin-counting rules and so on. This approach suppresses the production of heavier hadrons, and gives them a larger $\langle p_{\perp} \rangle$. Such a pattern is observed in data, and exists in the Gaussian approach mainly owing to particle decays, but there undershoots data.

Making the Gaussian p_{\perp} width, or temperature in case of the thermodynamical model, dependent on the close-packing of strings allows for modelling the influence of strings on each other in a simple way. An effective number of density of strings is introduced for low p_{\perp} 's, while high- p_{\perp} fragmentation tends to occur outside the close-packed string region and is left unaffected. Such a mechanism could e.g. be used to explain a changing flavour composition at high multiplicities.

Finally we implemented a simple model for hadronic rescattering, applied to the primary hadrons, before decays. The probability of two hadrons to rescatter is based on how close they are in phase space. By favouring a shift towards equal transverse velocities, it should also give higher $\langle p_{\perp} \rangle$ for heavy hadrons and lower for pions.

Not surprisingly we found the hadronic decays to limit the hoped-for effects. Specifically, most pions come from decays of heavier hadrons, and so the mechanisms intended to give less p_{\perp} to pions and more to kaons and protons are largely nullified. The mechanisms are also not simply additive; starting out from the thermodynamical model with its already-existing mass differentiation, the further effects of varying temperatures or hadronic rescattering are smaller than corresponding effects in the Gaussian approach.

Nevertheless the thermodynamical model is able to provide reasonable descriptions of observables such as the p_{\perp} spectrum of charged hadrons, the average transverse momentum as a function of the hadron mass, or the recently measured enhanced production of strange and multi-strange hadrons with increasing multiplicity. These observables have so far been described rather poorly by PYTHIA. And, given the small number of flavour parameters in the thermodynamical model, it is able to describe a reasonable number of e^+e^- data rather well, even if it can not compete with the many-more-parameter tunes of default PYTHIA.

To advance to the next level of sophistication within the line of research advocated here, it would be necessary to do a microscopic tracing of the full space-time evolution of the

event, both for partons and for hadrons, and including both production and decay vertices. This is nontrivial beyond the simple one-string picture, even in the cleaner e^+e^- events, and the further complications of MPIs and CR in hadronic events will make it even worse. What it would allow is a more detailed understanding of the close-packing both of strings and of hadrons. Combined with a more detailed modelling of hadronic rescattering, a more realistic picture may emerge.

Some of the limitations encountered here are likely still to remain, so further mechanisms may be at play, in addition to the ones studied here. This would not be the first time where a cocktail of smaller effects combine to give a significant signal. What is less likely is actually the opposite, that one single mechanism does it all. Specifically, whatever else may be going on, the close-packing of strings and hadrons appears unavoidable in high-multiplicity pp events, and collective-flow effects are here to stay. In sum, we have an interesting and challenging time ahead of us, where some of the most unexpected new LHC observations may well come in the low- p_\perp region rather than the in high- p_\perp one.

Acknowledgements

Helpful discussions with Peter Skands are gratefully acknowledged. Work supported by the MCnetITN FP7 Marie Curie Initial Training Network, contract PITN-GA-2012-315877. This project has also received funding in part by the Swedish Research Council, contract number 621-2013-4287, and in part from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 668679).

A Settings

Tab. 1 gives a list of the settings that have been changed with respect to default PYTHIA to obtain the results presented in sec. 4.2.

References

- [1] G. P. Salam, *QCD Theory Overview — Towards Precision at LHC*, presentation at LHCP 2016, Lund, Sweden, 2016.
- [2] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, *Parton Fragmentation and String Dynamics*, *Phys. Rept.* **97** (1983) 31–145.
- [3] R. D. Field and S. Wolfram, *A QCD Model for e^+e^- Annihilation*, *Nucl. Phys.* **B213** (1983) 65.
- [4] T. D. Gottschalk, *An improved description of hadronization in the QCD cluster model for e^+e^- annihilation*, *Nucl. Phys.* **B239** (1984) 349.
- [5] B. R. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*, *Nucl. Phys.* **B238** (1984) 492.

Table 1: PYTHIA 8 parameters and their values for tuning to LHC and LEP/SLC observables.

	LHC			LEP and SLC	
	Default	Gaussian	Thermal	Gaussian	Thermal
Switch to thermal model?					
StringPT:thermalModel	off	off	on	off	on
Gaussian width $\sigma = \sqrt{\kappa/\pi}$					
StringPT:sigma	0.335	0.33	-	0.33	-
σ prefactor for s quarks					
StringPT:widthPreStrange	1.0	1.2	-	1.2	-
σ prefactor for diquarks					
StringPT:widthPreDiquark	1.0	1.2	-	1.2	-
Fraction with enhanced σ					
StringPT:enhancedFraction	0.01	0.0	-	0.0	-
Temperature T					
StringPT:temperature	-	-	0.21	-	0.21
Baryon normalization factor					
StringFlav:MtoBratio	-	-	3.0	-	3.0
Suppression factor for s hadrons					
StringFlav:StrangeSuppression	-	-	0.5	-	0.5
r parameter in eq. (8) or (15)					
StringPT:expNSP	0.0	0.01	0.13	0.0	0.0
Range of MPI-based CR scheme					
ColourReconnection:range	1.8	1.8	1.1	-	-
Hadron rescattering (HR) on?					
HadronLevel:HadronScatter	off	on	on	off	off
HR with eq. (16) and (17)					
HadronScatter:mode	-	0	0	-	-
P_{ds}^{\max} parameter in eq. (17)					
HadronScatter:maxProbDS	-	0.25	0.5	-	-
MPI regularization parameter					
MultipartonInteractions:pT0Ref	2.28	2.34	2.5	-	-

- [6] A. Ortiz Velasquez, P. Christiansen, E. Cuautle Flores, I. Maldonado Cervantes, and G. Paic, *Color Reconnection and Flowlike Patterns in pp Collisions*, *Phys. Rev. Lett.* **111** (2013), no. 4 042001, [[arXiv:1303.6326](#)].
- [7] CMS Collaboration, V. Khachatryan et al., *Observation of Long-Range Near-Side Angular Correlations in Proton-Proton Collisions at the LHC*, *JHEP* **09** (2010) 091, [[arXiv:1009.4122](#)].

- [8] **ATLAS** Collaboration, G. Aad et al., *Observation of Long-Range Elliptic Azimuthal Anisotropies in $\sqrt{s} = 13$ and 2.76 TeV pp Collisions with the ATLAS Detector*, *Phys. Rev. Lett.* **116** (2016), no. 17 172301, [[arXiv:1509.4776](#)].
- [9] **CMS** Collaboration, V. Khachatryan et al., *Evidence for collectivity in pp collisions at the LHC*, [arXiv:1606.6198](#).
- [10] **ALICE** Collaboration, J. Adam et al., *Multiplicity-dependent enhancement of strange and multi-strange hadron production in proton-proton collisions at $\sqrt{s} = 7$ TeV*, [arXiv:1606.7424](#).
- [11] P. Braun-Munzinger and J. Stachel, *The quest for the quark-gluon plasma*, *Nature* **448** (2007) 302–309.
- [12] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner, *EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider*, *Phys. Rev.* **C92** (2015), no. 3 034906, [[arXiv:1306.0121](#)].
- [13] T. S. Biro, H. B. Nielsen, and J. Knoll, *Color Rope Model for Extreme Relativistic Heavy Ion Collisions*, *Nucl. Phys.* **B245** (1984) 449–468.
- [14] A. Bialas and W. Czyz, *Chromoelectric Flux Tubes and the Transverse Momentum Distribution in High-energy Nucleus-nucleus Collisions*, *Phys. Rev.* **D31** (1985) 198.
- [15] B. Andersson and P. A. Henning, *On the dynamics of a color rope: The Fragmentation of interacting strings and the longitudinal distributions*, *Nucl. Phys.* **B355** (1991) 82–105.
- [16] C. Bierlich, G. Gustafson, L. Lönnblad, and A. Tarasov, *Effects of Overlapping Strings in pp Collisions*, *JHEP* **03** (2015) 148, [[arXiv:1412.6259](#)].
- [17] P. Capiluppi, G. Giacomelli, A. M. Rossi, G. Vannini, and A. Bussiere, *Transverse momentum dependence in proton proton inclusive reactions at very high energies*, *Nucl. Phys.* **B70** (1974) 1–38.
- [18] R. Hagedorn, *Statistical thermodynamics of strong interactions at high-energies*, *Nuovo Cim. Suppl.* **3** (1965) 147–186.
- [19] R. Hagedorn, *Remarks on the thermodynamical model of strong interactions*, *Nucl. Phys.* **B24** (1970) 93–139.
- [20] R. Hagedorn, *Multiplicities, p_T Distributions and the Expected Hadron \rightarrow Quark - Gluon Phase Transition*, *Riv. Nuovo Cim.* **6N10** (1983) 1–50.
- [21] S. Barshay and Y. A. Chao, *Longitudinal mass, inclusive transverse-momentum distributions, proton proton elastic diffraction slopes and the total cross-section*, *Phys. Lett.* **B38** (1972) 225–228.
- [22] C. Tsallis, *Possible Generalization of Boltzmann-Gibbs Statistics*, *J. Statist. Phys.* **52** (1988) 479–487.

- [23] J. Cleymans, *On the Use of the Tsallis Distribution at LHC Energies*, in *16th International Conference on Strangeness in Quark Matter (SQM 2016) Berkeley, California, United States, June 27-July 1, 2016*, 2016. [arXiv:1609.0228](#).
- [24] A. A. Bylinkin and A. A. Rostovtsev, *Parametrization of the shape of hadron-production spectra in high-energy particle interactions*, *Phys. Atom. Nucl.* **75** (2012) 999–1005. [*Yad. Fiz.*75,1060(2012)].
- [25] T. Sjöstrand and M. van Zijl, *A Multiple Interaction Model for the Event Structure in Hadron Collisions*, *Phys. Rev.* **D36** (1987) 2019.
- [26] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026, [[hep-ph/0603175](#)].
- [27] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159–177, [[arXiv:1410.3012](#)].
- [28] X. Artru and G. Mennessier, *String model and multiproduction*, *Nucl. Phys.* **B70** (1974) 93–115.
- [29] B. Andersson, G. Gustafson, and B. Söderberg, *A Probability Measure on Parton and String States*, *Nucl. Phys.* **B264** (1986) 29.
- [30] G. 't Hooft, *A planar diagram theory for strong interactions*, *Nucl. Phys.* **B72** (1974) 461.
- [31] B. Andersson, G. Gustafson, and T. Sjöstrand, *How to Find the Gluon Jets in e^+e^- Annihilation*, *Phys. Lett.* **B94** (1980) 211–215.
- [32] **JADE** Collaboration, W. Bartel et al., *Experimental Study of Jets in electron - Positron Annihilation*, *Phys. Lett.* **B101** (1981) 129–134.
- [33] T. Sjöstrand, *Jet fragmentation of nearby partons*, *Nucl. Phys.* **B248** (1984) 469.
- [34] T. Sjöstrand and P. Z. Skands, *Baryon number violation and string topologies*, *Nucl. Phys.* **B659** (2003) 243, [[hep-ph/0212264](#)].
- [35] B. Andersson, G. Gustafson, and T. Sjöstrand, *On Soft Gluon Emission and the Transverse Momentum Properties of Final State Particles*, *Z. Phys.* **C12** (1982) 49.
- [36] F. E. Close, *An Introduction to Quarks and Partons*. 1979.
- [37] B. Andersson, G. Gustafson, and T. Sjöstrand, *A Model for Baryon Production in Quark and Gluon Jets*, *Nucl. Phys.* **B197** (1982) 45–54.
- [38] B. Andersson, G. Gustafson, and T. Sjöstrand, *Baryon Production in Jet Fragmentation and Upsilon-Decay*, *Phys. Scripta* **32** (1985) 574.
- [39] S. Plätzer and S. Gieseke, *Dipole Showers and Automated NLO Matching in Herwig++*, *Eur. Phys. J.* **C72** (2012) 2187, [[arXiv:1109.6256](#)].

- [40] P. Richardson and D. Winn, *Investigation of Monte Carlo Uncertainties on Higgs Boson searches using Jet Substructure*, *Eur. Phys. J.* **C72** (2012) 2178, [[arXiv:1207.0380](#)].
- [41] S. Chun and C. Buchanan, *A Simple plausible path from QCD to successful prediction of $e^+ e^-$ hadronization data*, *Phys. Rept.* **292** (1998) 239–317.
- [42] S. Abachi, C. Buchanan, A. Chien, S. Chun, and B. Hartfiel, *UCLA space-time area law model: A Persuasive foundation for hadronization*, *Eur. Phys. J.* **C49** (2007) 569–580, [[hep-ph/0612103](#)].
- [43] A. Buckley et al., *Rivet user manual*, [[arXiv:1003.0694](#)].
- [44] **STAR** Collaboration, B. I. Abelev et al., *Systematic Measurements of Identified Particle Spectra in pp , d^+ Au and Au+Au Collisions from STAR*, *Phys. Rev.* **C79** (2009) 034909, [[arXiv:0808.2041](#)].
- [45] **ALICE** Collaboration, B. B. Abelev et al., *Production of $\Sigma(1385)^\pm$ and $\Xi(1530)^0$ in proton-proton collisions at $\sqrt{s} = 7$ TeV*, *Eur. Phys. J.* **C75** (2015), no. 1 1, [[arXiv:1406.3206](#)].
- [46] **ATLAS** Collaboration, G. Aad et al., *Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC*, *New J. Phys.* **13** (2011) 053033, [[arXiv:1012.5104](#)].
- [47] **ALICE** Collaboration, L. Bianchi, *Strangeness production as a function of charged particle multiplicity in proton-proton collisions*, [[arXiv:1604.6736](#)].
- [48] **CMS** Collaboration, S. Chatrchyan et al., *Charged particle transverse momentum spectra in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV*, *JHEP* **08** (2011) 086, [[arXiv:1104.3547](#)].
- [49] **ALICE** Collaboration, J. Adam et al., *Pseudorapidity and transverse-momentum distributions of charged particles in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Phys. Lett.* **B753** (2016) 319–329, [[arXiv:1509.8734](#)].
- [50] **ALICE** Collaboration, J. Adam et al., *Measurement of pion, kaon and proton production in proton-proton collisions at $\sqrt{s} = 7$ TeV*, *Eur. Phys. J.* **C75** (2015), no. 5 226, [[arXiv:1504.0024](#)].
- [51] **CMS** Collaboration, V. Khachatryan et al., *Strange Particle Production in pp Collisions at $\sqrt{s} = 0.9$ and 7 TeV*, *JHEP* **05** (2011) 064, [[arXiv:1102.4282](#)].
- [52] **UA1** Collaboration, C. Albajar et al., *A Study of the General Characteristics of $p\bar{p}$ Collisions at $\sqrt{s} = 0.2$ -TeV to 0.9-TeV*, *Nucl. Phys.* **B335** (1990) 261–287.
- [53] T. Sjöstrand and V. A. Khoze, *On Color rearrangement in hadronic W^+W^- events*, *Z. Phys.* **C62** (1994) 281–310, [[hep-ph/9310242](#)].
- [54] J. R. Christiansen and P. Z. Skands, *String Formation Beyond Leading Colour*, *JHEP* **08** (2015) 003, [[arXiv:1505.0168](#)].

- [55] A. Chodos, R. L. Jaffe, K. Johnson, C. B. Thorn, and V. F. Weisskopf, *A New Extended Model of Hadrons*, *Phys. Rev.* **D9** (1974) 3471–3495.
- [56] **Particle Data Group** Collaboration, C. Amsler et al., *Review of Particle Physics*, *Phys. Lett.* **B667** (2008) 1–1340.
- [57] **British-Scandinavian** Collaboration, B. Alper et al., *Production Spectra of π^{+-} , K^{+-} , ρ^{+-} at Large Angles in Proton Proton Collisions in the CERN Intersecting Storage Rings*, *Nucl. Phys.* **B100** (1975) 237–290.
- [58] **British-Scandinavian-MIT** Collaboration, K. Guettler et al., *Inclusive Production of Low-Momentum Charged Pions at $x = 0$ at the CERN Intersecting Storage Rings*, *Phys. Lett.* **B64** (1976) 111–116.
- [59] G. Giacomelli and M. Jacob, *Physics at the CERN ISR*, *Phys. Rept.* **55** (1979) 1–132.
- [60] *Mathematica 11.0*. Wolfram Research, Inc., Champaign, Illinois, 2016.
- [61] M. Gyulassy and L. McLerran, *New forms of QCD matter discovered at RHIC*, *Nucl. Phys.* **A750** (2005) 30–63, [[nucl-th/0405013](#)].
- [62] U. Heinz and R. Snellings, *Collective flow and viscosity in relativistic heavy-ion collisions*, *Ann. Rev. Nucl. Part. Sci.* **63** (2013) 123–151, [[arXiv:1301.2826](#)].
- [63] G. Roland, K. Safarik, and P. Steinberg, *Heavy-ion collisions at the LHC*, *Prog. Part. Nucl. Phys.* **77** (2014) 70–127.
- [64] R. D. Field and R. P. Feynman, *A parametrization of the properties of quark jets*, *Nucl. Phys.* **B136** (1978) 1.
- [65] **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** (2010) 022002, [[arXiv:1005.3299](#)].
- [66] **SLD** Collaboration, K. Abe et al., *Production of π^+ , π^- , K^+ , K^- , p and anti- p in light (uds), c and b jets from Z^0 decays*, *Phys. Rev.* **D69** (2004) 072003, [[hep-ex/0310017](#)].
- [67] **ALEPH** Collaboration, R. Barate et al., *Studies of quantum chromodynamics with the ALEPH detector*, *Phys. Rept.* **294** (1998) 1–165.