

Closepacking effects on strangeness and baryon production at the LHC

J. Altmann^{a,*}, L. Bernardinis^b, P. Skands^a and V. Zaccolo^b

^a*School of Physics and Astronomy, Monash University,
Wellington Road, Clayton, VIC-3800, Australia*

^b*Dipartimento di Fisica dell'Università and Sezione INFN,
Via Alfonso Valerio 2, Trieste, 34127, Italy*

E-mail: javira.altmann@monash.edu, lorenzo.bernardinis@phd.units.it,
peter.skands@monash.edu, valentina.zaccolo@units.it

ABSTRACT: Data from the LHC show a rise in strange-hadron production with charged-particle multiplicity in pp collisions. The Monte-Carlo event generator PYTHIA, using its default Monash tune, instead predicts constant strangeness. We investigate a mechanism invoked during hadronization called “string closepacking”, where overlapping strings generate a background field, here assumed to be predominantly aligned with the beam axis. This increases the effective string tension, reducing strangeness suppression and thus enhancing strangeness production. We tune this model to LHC data and contrast it with several alternatives. We comment specifically on the challenge of simultaneously describing the non-strange p/π ratio, and introduce a mechanism which may act to suppress this. Many of the salient particle-production ratios can be qualitatively described by this model, although the Ξ_c/D ratio and the shape of $p_\perp$ spectra remain challenging to account for.

KEYWORDS: Specific QCD Phenomenology, Jets and Jet Substructure

ARXIV EPRINT: [2512.00671](https://arxiv.org/abs/2512.00671)

*Corresponding author.

Contents

1	Introduction	1
2	Theoretical modelling	4
2.1	Basics of the Lund string model	4
2.2	Closepacking	8
2.3	Strange junctions	16
2.4	Popcorn destructive interference	17
3	Constraints and results	19
3.1	Models for comparison	20
3.2	Tuning methodology	21
3.3	Tuning outcomes	28
3.4	Heavy flavour	31
4	Conclusions and outlook	32
A	Diquark scaling for closepacking	35
B	Table of Casimir-scaling coefficients	36
C	Reference tune parameters	37
D	Further plots	38
E	Colour map for goodness-of-fit assessment	40
F	Tuning observables	40
F.1	Tuning step 1	41
F.2	Tuning step 2	41
F.3	Tuning step 3	42

1 Introduction

The Lund string model for hadronization [1], as implemented in the Monte Carlo event generator PYTHIA [2, 3], has been shown to well describe data from e^+e^- colliders such as LEP and SLC [4–14] as well as having had some success in describing pp collisions [11, 14–17]. Recent data from the LHC however has highlighted discrepancies between PYTHIA predictions and pp collision events, prompting further investigations into string modelling, particularly in small systems.

The baseline Lund string model (LSM) describes the fragmentation of a string of arbitrary invariant mass $\gg \Lambda_{\text{QCD}}$. The string is taken to be in vacuum and isolated from any (other) sources of colour fields. This applies especially well to the context of hadronic Z decays, which is also the main reference case for constraining the nonperturbative modelling parameters

of the LSM. The main generalization when going to pp collisions is that multiple parton-parton interactions (MPI) [18, 19] as well as the colour-octet nature of gluon exchanges intrinsically lead to a picture of multiple such strings fragmenting simultaneously. Both the number and the lengths (invariant masses) of these strings can vary significantly from event to event, and when this is furthermore allowed to depend upon the proton-proton impact parameter [20], one can successfully account [18] for the very wide ($\sim$ negative-binomial) multiplicity distributions that are observed in minimum-bias collisions [21], with large long-range correlations [22, 23], as well as for the comparatively large underlying events with a hard trigger (pedestal effect) [24–30].

It is worth pausing to note that, in the context of the baseline LSM, these successes can be achieved while remaining within the comparatively simple paradigm that the fragmentation of a system of multiple strings can be described as the summed fragmentations of each of its constituent strings, with universal parameters — taken to be the same as those measured in hadronic Z decays.

This, however, leads to predictions, e.g., of approximately universal ratios of the production of each hadron spin and flavour type relative to the total amount of hadron production. Such universal behaviours have been convincingly ruled out by more recent measurements, in particular by the ALICE experiment [31–34] whose excellent particle identification (PID) capabilities down to very low track momenta have played a crucial role in reconstructing an almost complete set of ground-state ($L = 0$) hadrons.¹ We also note that analogous and sometimes even more striking conclusions are reached in the heavy-flavour sector [35].

This has raised questions about whether hadronization in pp collisions, especially at high particle densities, differs in some more fundamental way from hadronization in Z decays. If one wishes to remain within a string context, one relatively obvious possibility (with some benefit of hindsight) is that it is the above-mentioned assumption of independently fragmenting strings which is violated.

This question of so-called collective effects is one that has already been posed and explored in models such as Rope hadronization [36, 37] and string shoving [38–40], which model effects on the string tension and string kinematics respectively. An interesting complementary proposal is that, in a high-density “core” of high-multiplicity pp collisions, the string picture may break down and small droplets of quark-gluon plasma (QGP) be formed instead, while strings continue to fragment more or less as in vacuum in a low-density “corona” [41–43].

Interest in collective nonperturbative effects has recently been reignited by the above-mentioned measurements of clear breakdowns of universality, such as the ratios of strange hadrons to pions from the ALICE collaboration [32, 33], which show an increase in strange hadron production relative to charged multiplicity in pp collisions. This rise in strangeness gives a good indication that nonperturbative strange-quark production (e.g., via string breaks) is altered in correlation with overall particle production, an effect which the baseline LSM fails to capture.

¹In the sense of having measurements of at least one isospin state for almost every distinct spin and strangeness combination, sometimes more. The main exception is the Δ family of baryons whose strong decays make them particularly challenging to reconstruct.

The existing Rope hadronization model has shown promising success in describing such strangeness-enhancement effects, doing so by implementing string-tension modifications as an extension to the baseline LSM. More specifically, the Rope model increases the string tension in higher-density string systems, resulting in a reduced strangeness suppression (i.e., strangeness enhancement) in string breaks. The Rope hadronization model construction however requires spacetime modelling of string breaks, which is computationally expensive and in contrast to the standard LSM fragmentation procedure which is implemented entirely in momentum space. As an alternative, we update and extend a simpler momentum-space model for string-tension modifications, called “closepacking” which was first proposed in the context of a thermal string-breaking model [44]; in this work, we generalize the implementation to Schwinger-type string breaking and introduce sensitivity to relative flux orientations on the basis of Casimir scaling. The model employs the same basic premise of an increased string tension as the more elaborate Rope model and hence should share a number of its features. But where the Rope model attempts to trace the explicit evolution of the fragmenting parton systems in spacetime, with multi-string ropes breaking strand by strand, the closepacking model currently only attempts to describe a time-averaged “effective background” for each string, as a function of rapidity along the beam axis. The main consequence is a rapidity-dependent effective tension that depends on each string’s environment (relative to the beam axis).

As such, our closepacking model is currently only applicable to pp collisions and not to heavy-ion ones where spacetime distances between string systems can be large compared to a typical hadron size. Such systems would be better addressed by the Rope model or by QGP-inspired models such as EPOS [42, 43]. In its current formulation, the closepacking model can also not be applied to systems like $e^+e^- \rightarrow W^+W^- \rightarrow$ hadrons where the beam axis plays no special role, nor will it produce any strangeness-enhancement effects among high- $p_\perp$ hadrons inside jets. We plan to lift some of these restrictions in future work but for the time being focus on minimum-bias pp collisions where effects along the beam axis should dominate.

Interestingly, though strangeness-enhancement modelling has previously been studied by the Rope hadronization model, no models of string collectivity have thus far attempted to address the overprediction of the proton-to-pion ratio observed by ALICE in pp collisions [32]. The default Monash tune [11], despite having its diquark production probabilities tuned to LEP data, overpredict the proton-to-pion ratio seen by ALICE by around 20%, prompting the question of how the p/π ratio appears lower at the LHC than at LEP and thus hinting at a flaw in the modelling of baryon production in denser string environments. This overprediction only becomes further exacerbated with the inclusion of diquark enhancement via colour ropes or so-called junction baryons in the QCD Colour Reconnection (CR) model [45]. Production of these junction baryons via so-called string junctions [46, 47] has been shown to be important in describing increased baryon-to-meson ratios at LHC compared to LEP, such as the ratio $(\Lambda + \bar{\Lambda})/2K_S^0$ (0.191 ± 0.004 at 91 GeV LEP [48], and 0.267 ± 0.002 at 7 TeV LHC [49]), and the heavy-flavoured $\Lambda_c^\pm/D^0$ (0.130 ± 0.013 at 91 GeV LEP [48], and 0.440 ± 0.026 at 13 TeV LHC [50]). Given the apparent importance of junction baryons, this begs the question of whether modelling of diquark production is to blame for the overpredicted p/π ratio. Building upon the existing popcorn mechanism [51, 52] (which provides a more sophisticated model for diquark production than what is used by the baseline string model), in this paper we

introduce a novel mechanism to decrease diquark formation in denser string environments and thus reduce proton production in pp collisions.

The remainder of the paper will be structured as follows; first we will provide a brief overview of the Lund string model in section 2.1, with particular focus on modelling of flavour and baryon production. Sections 2.2, 2.3, and 2.4 explore the theory behind the newly implemented models, with section 2.2 focusing on the closepacking model, section 2.3 the so-called “strange junctions” model, and section 2.4 introduces the novel popcorn destructive interference model. Following this, in section 3 we will provide a more detailed outline of the aforementioned existing models which we shall use for comparisons, and highlight the key relevant PYTHIA parameters. Then we provide a first tuning effort implementing the new models introduced in section 2, with comparisons focusing on results from the ALICE collaboration [32, 33]. Proceeding the tuning efforts, we briefly explore the implications on the heavy-flavour sector.

For ease of reference, throughout the text technical parameters in the code are emphasized with `typewriter` font, using `Violet` for existing PYTHIA parameters while new ones introduced in this work are highlighted in `Burgundy`.

2 Theoretical modelling

2.1 Basics of the Lund string model

The Lund String model (LSM) is a semi-classical hadronization model based on QCD confinement and relativistic string dynamics. Lattice QCD results [53] show a linear potential between static colour charges at non-perturbative scales, which can be parameterized phenomenologically, e.g., by the Cornell potential [54]. Such a linear potential is equivalently interpreted as a string with a constant tension, thus allowing the colour confinement field to be represented as a 1+1 dimensional string obeying relativistic string dynamics [55] with characteristic string tension $\kappa \sim 1 \text{ GeV/fm}$. These confining potentials — or so-called Lund strings — are expected to form between colour-connected charges, i.e., colour charges that form overall colour-singlet states.

In high-energy collision processes such as those at the LHC, these colour-connected partons move apart from one another at high energies, in turn stretching the colour fields. Given a sufficiently large confining potential, it can become energetically favourable to break the confinement field and spontaneously pair create a quark-antiquark (or diquark-antidiquark) from the vacuum. This pair creation process from string breaking is what we call Lund string fragmentation, and is used to model the hadronization process. A given string will break — or “fragment” — till there is no longer sufficient energy to keep breaking the string, resulting in observable final-state hadrons. Figure 1 shows a simple schematic depiction of hadronization given the string breaking picture described above.

To model the (di)quark pair creation process from string breaks, quantum tunnelling analogous to the Schwinger mechanism for QED [56] is used. The Schwinger mechanism in QED is used to model e^+e^- spontaneous pair creation in the presence of a strong electric field in a vacuum, with the tunnelling probability following a Gaussian distribution in transverse

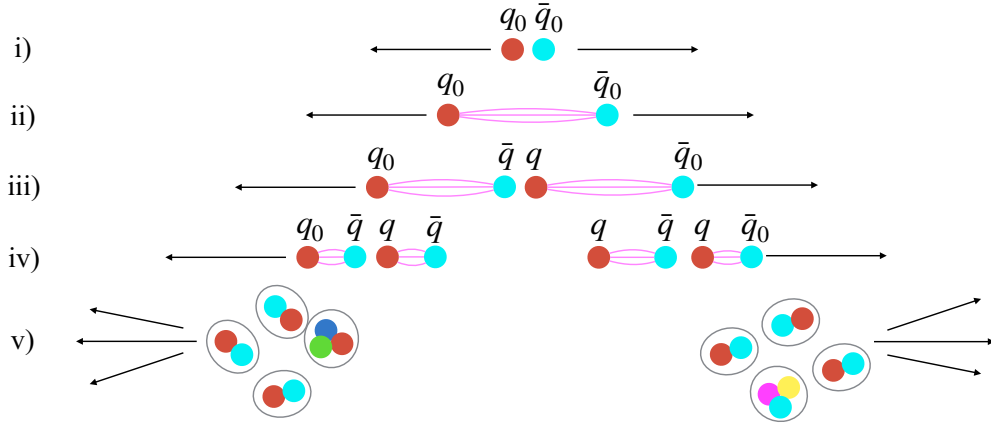


Figure 1. Schematic illustration of hadronization given a string representation of the colour confinement field. (i) The initial red-antired quark-antiquark pair, $q_0 \bar{q}_0$, moving apart at high energies. (ii) The string forms between the $q_0 \bar{q}_0$ pair. (iii) The initial string breaks and a red-antired $q \bar{q}$ pair is created at the site of the string break. (iv) Further string breaking occurs. (v) Final jet of hadrons formed.

mass. The analogous QCD tunnelling probability used by the baseline LSM is given by,

$$\exp\left(\frac{-\pi m_q^2}{\kappa}\right) \exp\left(\frac{-\pi p_{\perp q}^2}{\kappa}\right) = \exp\left(\frac{-\pi m_{\perp q}^2}{\kappa}\right), \quad (2.1)$$

where $m_{\perp q}$ is the transverse quark mass, $p_{\perp q}$ is the transverse momentum, and m_q is the effective quark mass. Given QCD is a non-Abelian SU(3) gauge theory in contrast to the Abelian QED, the Schwinger mechanism here serves only as an analogy. Exact solutions have been explored, e.g., in [57].

The transverse mass dependence of the Schwinger mechanism leads to a Gaussian transverse momentum distribution along with a suppression of larger mass quark flavours. Effectively this leads to reduced production rates of strange quarks and diquarks relative to up/down quark production. Note that given a string tension of $\kappa \sim 1 \text{ GeV/fm}$, string breaks can only pair-create light-flavoured (di)quarks, i.e., up-, down-, or strange-flavoured. Charm and bottom quark production via string breaks would be heavily suppressed in the Schwinger picture ($\lesssim 10^{-11}$), and thus in PYTHIA their production probabilities (in string breaks) are set to zero. This agrees with the expectation that heavy quark flavours (and high- $p_{\perp}$ excitations) should only be produced via perturbative processes, and not from non-perturbative mechanisms such as string breaks.

Due to the ambiguity around both the quark masses and the precision of the string tension, flavour selection in PYTHIA does not directly implement mass-suppression factors according to the Schwinger mechanism Gaussian. Instead, the mass suppression is parameterized in terms of a tuneable ratio of probabilities, i.e., `StringFlav:probStoUD` ($P(s : u/d)$) and `StringFlav:probQQtoQ` ($P(qq : q)$), where the former is the probability of strange quark production relative to up/down production, and the latter is the diquark-to-quark probability.

Important to note is that the form of eq. (2.1) means that an increase in the string tension value will have two key consequences on the tunnelling probabilities — a broadening of the

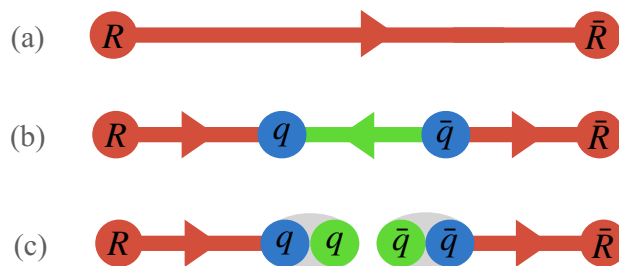


Figure 2. An illustration of the stepwise process of the popcorn mechanism given a red-antired string. (a) the unbroken red-antired string. (b) a blue-antiblue virtual colour fluctuation on the red-antired string, which results in a short segment of “green-coloured” string. (c) a green-antigreen quark pair creation that combined with the blue-antiblue quark fluctuations forms a red-antired diquark-antidiquark pair, breaking the string.

$p_\perp$ spectrum, and a reduction of the mass suppression. This relationship between string tension and flavour will be foundational for the closepacking mechanism in section 2.2.

Mechanisms for baryon production: diquarks and junctions. Within the LSM, there are two primary mechanisms for baryon production; diquark-antidiquark pair production via string breaks, and junction string topologies.

In the baseline LSM, diquark-antidiquark pair production is modelled on equal footing as quark-antiquark production — i.e., spontaneously pair created out of the vacuum as a triplet-antitriplet pair. This means that the mass suppression of diquark production is directly governed by the diquark masses and the Schwinger mechanism.

An alternative interpretation of diquark production is the so-called “popcorn mechanism” [51, 52]. Instead of considering diquarks as directly tunnelling from the vacuum as constituent particles, the popcorn mechanism supposes that diquarks are formed via the combination of two successive colour fluctuations on a string. An example demonstrating this colour-fluctuation based argument for diquark formation can be seen in figure 2. Per this example, consider a red-antired $q\bar{q}$ pair with a red confining field between them (figure 2(a)). This red confining field is characterized by two defining features — its red colour, and by its “flux” direction as indicated by an arrow on the string. Here the term “flux” refers to the direction of the colour flow, conventionally orientated as flowing from the triplet to the antitriplet colour charge. Suppose then we allow arbitrarily coloured fluctuations to form on this red string, e.g. a blue-antiblue coloured $q\bar{q}$ fluctuation. Unlike conventional string breaks which require the tunnelling of a red-antired $q\bar{q}$ pair to break the confining field, a virtual blue-antiblue pair would not initially “break” the string as there would remain a non-zero colour field between the blue-antiblue pair. More specifically, the combination of the red and blue colour charges results in an overall antigreen charge, thus producing a green colour field oriented in the opposite flux direction to the initial red string as can be seen in figure 2(b). From this point, tunnelling of a green-antigreen $q\bar{q}$ pair is able to break the green string segment, resulting in the formation of a blue-green diquark and an antiblue-antigreen antidiquark as shown in figure 2(c).

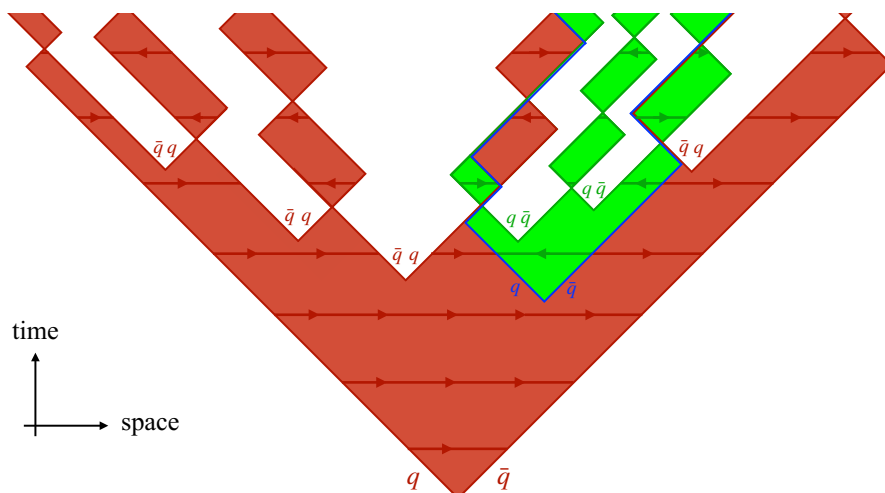


Figure 3. Minkowski diagram of the popcorn mechanism showing diquark-antidiquark formation and a single meson produced between the baryon and antibaryon i.e., $BM\bar{B}$ configuration. The flux direction of the strings are indicated by arrows, with colour flow from the triplet to antitriplet colour charge. In this picture, we assume the green-antigreen fluctuation forms with zero energy and momentum.

A key consequence of modelling diquark formation using the “popcorn mechanism” is the possibility of meson production occurring between diquark-antidiquark pairs. After an initial colour fluctuation on a given string, rather than a single $q\bar{q}$ pair breaking the intermediate string per figure 2, suppose multiple colour fluctuations were allowed. Given the same colour structure as the example in figure 2, should we allow for multiple green-antigreen fluctuations to occur on the intermediate green string, it would result in the formation of one or more green-antigreen mesons between the resulting diquark-antidiquark pair. Figure 3 depicts an example of this process in the form of a Minkowski space-time diagram. Consequently this reduces the local correlation of the baryon-antibaryon pair formed, as it allows $BM\bar{B}$ configurations alongside $B\bar{B}$ (where B and M denotes baryons and mesons respectively). Contrastingly, in the standard string breaking picture diquark-antidiquark pairs are always created next to one another at string breaks, and hence only $B\bar{B}$ configurations are permitted.

In PYTHIA, these successive colour fluctuations used by the popcorn mechanism are not explicitly dynamically modelled. Rather, an additional parameter `StringFlav:popcornRate` is introduced which represents the probability of the $BM\bar{B}$ configuration occurring relative to the $B\bar{B}$ configuration (with configurations with multiple intermediate mesons — $BMM\bar{B}$, $BM^3\bar{B}$, etc. — neglected as they are expected to be suppressed). Note that the diquark production probability is assumed to be constant along a string regardless of whether one uses the popcorn or Schwinger mechanism for diquark production, and hence diquark production remains set by the tuneable parameter `StringFlav:probQQtoQ`.

A reexamination of diquark production via the popcorn mechanism given high density string systems will be the focus of section 2.4 of this paper, with a particular focus on the effect on the p/π ratio.

Additional to diquark-antidiquark pair production, baryon formation can also occur via so-called string junctions. Thus far, the representations of strings depicted above have

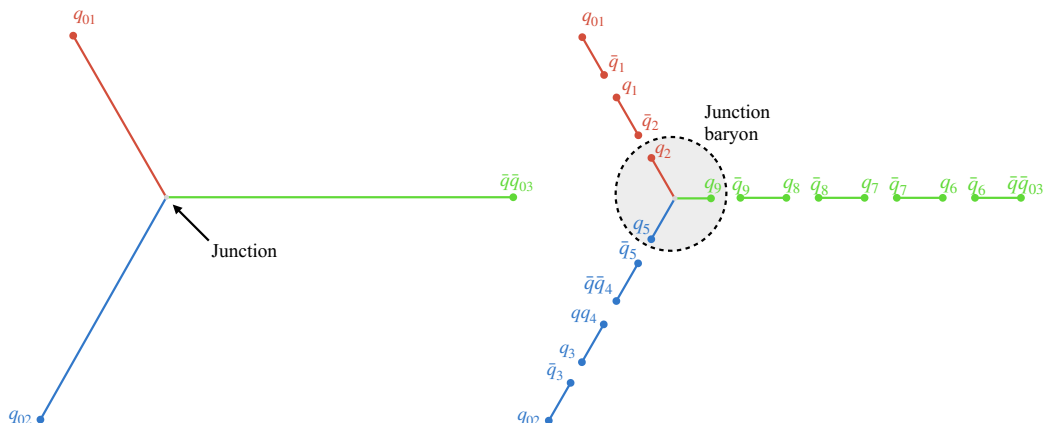


Figure 4. Junction configuration before (left) and after (right) fragmentation, with quarks $q_2 q_9 q_5$ together forming the so-called junction baryon.

all been “dipole” string topologies, defined by a string spanning between a triplet and antitriplet endpoint forming the overall colour-anticolour singlet state. However given an $SU(3)$ representation of colour, one should also consider the confining potential formed due to the red-green-blue colour singlet — a junction string [46, 47]. Such junction configurations are typically represented by a Y-shaped string as shown in the left image of figure 4, consisting of three junction “legs” with the “junction” itself being the meeting point of the three legs. Importantly, the baryon number of junction configurations are conserved, meaning that after fragmentation, the quarks surrounding the (anti)junction itself form a (anti)baryon. An example of a junction topology before and after fragmentation is shown in the left and right images of figure 4 respectively. For more detail on the modelling of such junctions and the fragmentation procedure, see [47].

Though the field towards the endpoints of each junction leg would be expected to behave similar to a dipole string, there remains ambiguity about field properties directly around the junction itself. For example, one may ask whether the energy density around a junction should be treated the same as a dipole string in vacuum. Such a question and implementation of this ambiguity will be further explored in section 2.3.

2.2 Closepacking

The closepacking model was first introduced in an earlier study in the context of thermodynamical string breaks [44]. In this paper, we generalize the model to conventional (Schwinger-type) string fragmentation, introduce sensitivity to relative flux orientations, and explore a few further optional implementation modifications.

Closepacking describes the collective effect of multiple strings in the near vicinity to one another. Typically in the LSM, the confining potential is approximated as a 1+1 dimensional field line, which fragments independently. However, realistically one should treat them more as flux tubes with non-zero thickness. Analogous to parallel conducting wires exerting an influence on one another — or perhaps more accurately vortex lines in a superconductor or superfluid — one would expect nearby strings to also affect each other. We suppose this background of surrounding strings manifests itself as an altered energy density, i.e., an

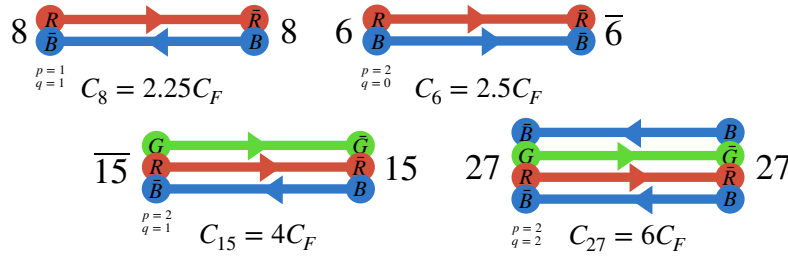


Figure 5. Examples of QCD multiplets in overall colour-singlet states and their colour factors according to Casimir scaling. The total number of parallel and antiparallel strings, labelled by p and q respectively, correspond to the number of free indices on the left-hand “side” of the multi-string systems that are contracted with those on the right-hand side, with $p > q$. See also appendix B.

effective enhanced string tension, κ_{eff} . Given the Gaussian form of the Schwinger mechanism, eq. (2.1), an increased effective string tension would correspond to reduced mass suppression of the quark-antiquark pairs produced in string breaks, and thus would lead to enhanced strange-quark production.

To motivate how the effective string tension should scale with the density of surrounding strings, we start by following in the footsteps of the Rope model, noting that the “total colour charge” of an irreducible representation of SU(3) with p fundamental (colour) indices and q antifundamental (anticolour) ones is given by the quadratic Casimir operator,

$$C_2(p, q) = \frac{p^2 + pq + q^2 + 3p + 3q}{3}. \quad (2.2)$$

For example, a quark (which has one free colour index) has $C_2(1, 0) = C_F = 4/3$, while a gluon (which has one free colour index and one free anticolour one) has $C_2(1, 1) = C_A = 9/3 = 3$. For reference, we give explicit colour factors for irreducible SU(3) representations with up to 8 free indices in appendix B. Assuming that a preceding stage of colour reconnections has acted to reduce (screen) colour charges so that only ones that cannot be further screened remain, the assumption that the colour charges defining the hadronizing string configurations constitute irreducible representations seems reasonable, at least in a first approximation. The assumption that the linear part of the QCD potential — and hence the total effective string tension — scales with this effective colour charge is referred to as Casimir scaling. We note that lattice studies support this assumption, though still with relatively large error bars [58].

Since the fundamental (quark-antiquark) string constitutes the baseline case for the LSM, it is useful to parameterize the colour factors of higher QCD representations with respect to that of the triplet, $C_F = 4/3$. Figure 5 provides a simple pictorial representation of some examples of such multiplets with their corresponding Casimir-scaled colour factors relative to C_F .

Let us suppose the net tension of the colour system scales proportional to these colour factors. Assuming that each string in a given multiplet shares an equal contribution to this modified colour factor — i.e., we should scale each strings tension by a factor of $(C_R/C_F)/(p+q)$ — one can then simply parameterize the effective string tension in terms of the p and q values,

$$\kappa_{\text{eff}} = \frac{C_R}{(p+q)C_F} \kappa_0 = (1 + 0.25p + 0.125q) \kappa_0. \quad (2.3)$$

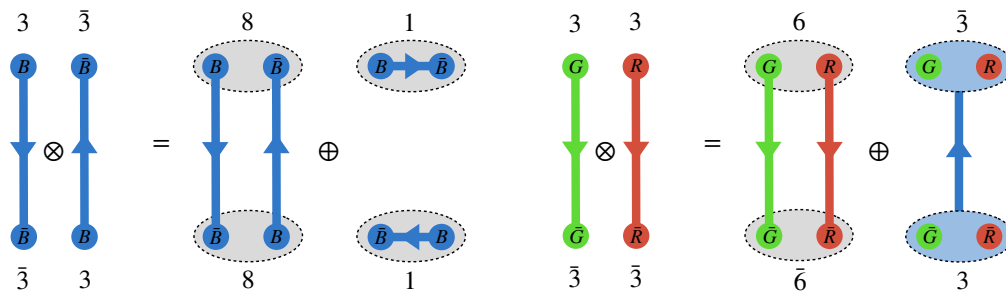


Figure 6. Pictorial representations possible multiplet representations for the case of two nearby strings, as dictated by SU(3) algebra. Left image: $3 \otimes \bar{3} = 8 \oplus 1$. Right image: $3 \otimes 3 = 6 \oplus \bar{3}$.

For the purposes of calculating the multiplet representation, R , for a given string, the “flux orientation” will refer to the direction relative to the string in question being fragmented. To make this clear, take the example of an octet configuration with colour factor $C_8 = C_A = 2.25C_F$. As the octet is comprised of two strings, each will contribute a factor of $1.125C_F$, and hence we assume the effective tension for a string in an octet configuration is given by $1.125\kappa_0$, making the total tension (summed over the two strings) $2.25\kappa_0 = (C_A/C_F)\kappa_0$. See table 5 in appendix B for further examples.

Important to note is that the colour-factor calculation with p and q here assumes the collective of strings is in the highest multiplet representation. More specifically, according to SU(3) algebra the multiplet representation of a parallel and an antiparallel string (i.e., $p = 1, q = 1$) does not necessarily always result in an octet string configuration. This is as SU(3) algebra dictates that there is a $1/9$ probability of a colour and anticolour charge in combination “screening” each other and resulting in a singlet state, i.e., $3 \otimes \bar{3} = 8 \oplus 1$. What this translates to in terms of string configurations can be seen in figure 6. As mentioned above, we work under the assumption that QCD CR [45], which proceeds the fragmentation procedure, effectively models any screening effects such that p and q post CR have already been reduced to their smallest irreducible-representation equivalents.

Though Casimir scaling provides a well-motivated foundation for string-tension modifications for pointlike static colour charges, we must also consider how this relates to the dynamical environment of collision systems, in which the total effective colour charges are made up of triplet quarks and octet gluons whose spatial separations are changing at or near light speed. In small systems like hadronic e^+e^- collisions and pp collisions, the individual colour charges originate from a space-time region that can be taken to be small compared with a typical hadron size. Taking a point-like origin as our starting point (which excludes heavy-ion collisions from consideration in the present formulation of the model), the main effect we shall seek to account for is the fact that the individual fragmenting strings do not sit right on top of each other (as in the lattice case) but instead will be expanding along axes that are determined by their individual momenta.

In this work, we shall not seek to account for the relative orientations (and boosts) of strings in detail. Instead, we allow for dampening factors that reduce the effect of the baseline Casimir scaling, to reflect that:

1. on average, strings will be some distance apart by the time they fragment;
2. high- $p_{\perp}$ hadrons are likely to be produced in spacetime regions well away from large string overlaps.

The first change we implement by generalizing eq. (2.3) with the introduction two adjustable constants,

$$\frac{\kappa_{\text{eff}}}{\kappa_0} = 1 + c_p (p + \omega q) , \quad (2.4)$$

where c_p now governs the overall (average) closepacking strength and $\omega \in [0, 1]$ governs the relative strength of contributions from strings with flux orientations antiparallel to the one being fragmented. Full-strength Casimir scaling corresponds to $c_p = 0.25$ and $\omega = 0.5$.

In reality, we expect $c_p < 0.25$ since, as discussed above, strings presumably fragment some distance apart, so that 0.25 should be considered a limiting maximum value (at least within the paradigm of Casimir scaling). A lower effective strength parameter may also account for imperfections in the colour-reconnection modelling leaving some representations in the modelling which could have been further reduced. Since we expect closepacking to impact both the flavour (strange and diquark production rates) and $p_{\perp}$ distributions of string breaks (detailed discussions on this follow below), in the PYTHIA implementation we allow for the strength of closepacking effects on each of these distributions to be set independently. This is done by allowing three separate c_p values, one for each corresponding physics effect,

$$c_p = \begin{cases} \text{ClosePacking:enhanceStrange} \\ \text{ClosePacking:enhanceDiquark} \\ \text{ClosePacking:enhancePT} \end{cases} \quad (2.5)$$

For ω , the PYTHIA implementation uses a single universal value,

$$\omega = \text{ClosePacking:fluxRatio} . \quad (2.6)$$

As mentioned above, the ratio for Casimir scaling is $\omega = 0.5$. The old closepacking model [44] which was insensitive to the flux direction would correspond to using $\omega = 1$, while for $\omega = 0$ only strings with flux directions parallel to each other would be allowed to increase each others' effective string tension, while strings with fluxes antiparallel to each other would be insensitive to each other. We believe this range of options gives a reasonable amount of flexibility to explore the modelling space.

The next question one must address is how to measure the multiplet representation a string is in given the distribution of colour charges in collision events is not comprised of a neat collection of (anti)parallel strings with well-defined colour-multiplet endpoints as in figure 5. This means that we need some measure of the number of nearby surrounding strings, alongside a way to take into account partial string overlaps. To do so, in the following we use rapidity with respect to the beam axis to determine whether two strings overlap. This variable choice is a logical approximation for pp collisions as the majority of strings are expected to be oriented along the beam axis due to beam remnants and low- $p_{\perp}$ multi-parton interactions. Given this rapidity measure for calculating p and q , and under the assumption that QCD CR

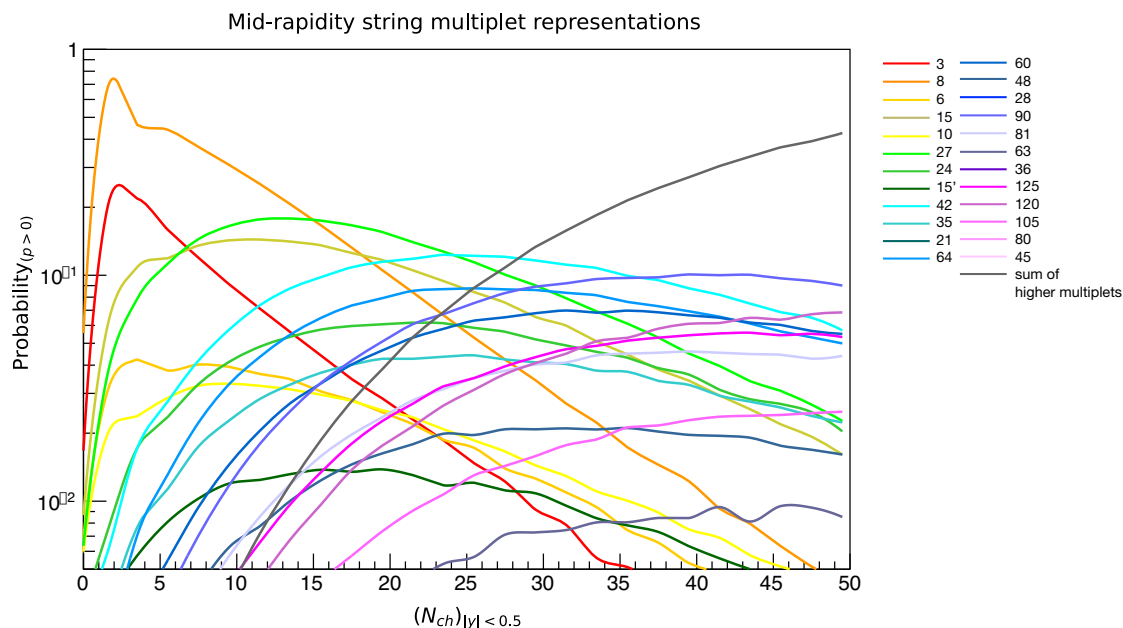


Figure 7. The probability of colour multiplet representations at $y=0$ with respect to the midrapidity charged multiplicity for 7 TeV inelastic events using QCD CR ($m_\tau = 2$) [45] parameters. Only events with $p > 0$ at $y = 0$ have been included. Flux configurations and Casimir-scaling factors for each multiplet are given in table 5. The highest multiplet shown here, the **125**, can be reached with just 4 gluon exchanges, if none of their colour charges screen each other.

takes into account any screening effects, figure 7 shows the likelihood of different multiplet representations at midrapidity in 7 TeV inelastic pp collisions, in fully simulated PYTHIA events using the QCD CR model (with time-dilation mode $m_\tau = 2$). For high-multiplicity events, we see that very high multiplets dominate, meaning that high-multiplicity events correspond to dense string systems (as one would reasonably expect), and one would hence expect closepacking effects to be particularly significant at these multiplicities. Note however that using rapidity with respect to the beam axis to measure string overlaps would no longer hold as a good approximation in hadronic e^+e^- events, for which the beam axis plays no special role; alternative approaches allowing for overlaps along generic axes will be explored in future studies.

To take into account partial string overlaps, the effective multiplet representation is determined on a per-string-break basis — i.e., using the rapidity associated with a given string break and comparing to the rapidity extents of the other strings in the collision system. This begs the question of how to determine the rapidity of a given string break. For this, we use a simple approximation of the rapidity of the produced hadron rather than the rapidity of the string break itself.² The flavour and transverse momentum of the produced hadron however are themselves dependent on the string tension (details of the dependence outlined in following sections), and thus sensitive to the number of surrounding strings. To

²This, again, is an approximation that could be lifted in a future study, but it would require solving the string world-sheet equations for a varying tension, which would be a nontrivial research project in its own right.

navigate around this issue, we follow the original closepacking implementation [44] and create a trial hadron with an average hadronic mass and transverse-momentum selection given the standard vacuum string-tension probabilities. The rapidity of this trial hadron is then used to determine the number of overlapping strings at that rapidity, which is thence used to calculate κ_{eff} to be used for the actual string break.

To further capture the dynamical nature of collisions, a concept of the separation of strings needs to be introduced. Under the assumption that we are working with hadron collisions, the bulk of strings will be aligned predominantly along the beam axis. However, hard interactions may also give rise to high- $p_{\perp}$ jets. The leading partons in such jets will move far from the beam axis before fragmenting and thus we don't expect them to experience as much closepacking tension enhancement as partons formed near the beam axis. Conversely, partons formed near the beam axis will be nearest the bulk collective of strings and thus experience maximal closepacking effects. To implement this effect, we use the $p_{\perp}$ of the corresponding hadrons as a measure of whether that hadron was produced “far away” from the beam axis ($p_{\perp} \gg \mathcal{O}(\Lambda_{\text{QCD}})$) or “near” the beam axis ($p_{\perp} \sim \mathcal{O}(\Lambda_{\text{QCD}})$). To do so, we require the functional form for κ_{eff} to scale such that when $p_{\perp} \rightarrow \infty$, the tension returns to the vacuum value (i.e., $\kappa_{\text{eff}} \rightarrow \kappa_0$), and when $p_{\perp} \rightarrow 0$ there are maximal closepacking effects. We choose a simple functional dependence inversely proportional to $1 + p_{\perp}^2/p_{\perp 0}^2$, with regularization parameter $p_{\perp 0}$ (**ClosePacking:PT0**), and similarly to determining rapidity overlaps, the $p_{\perp}$ used here is that of the trial hadron, $p_{\perp \text{Had}}$. This allows very soft hadrons produced in such jets to still “feel” collective effects, while the leading hadrons will not. The use of $p_{\perp}$ here is admittedly not perfect and can be challenged. Nevertheless, it allows us to interpolate between “soft” fragmentation aligned with the beam axis and “hard” fragmentation (along jet axes), the latter of which we presume still takes place in a good approximation to vacuum.

This gives us our final, fully generalized effective string-tension formula,

$$\kappa_{\text{eff}} = \left(1 + c_p \left(\frac{p + \omega q}{1 + p_{\perp \text{Had}}^2/p_{\perp 0}^2} \right) \right)^{2r} \kappa_0, \quad (2.7)$$

with we remind that exact (unsuppressed) Casimir scaling corresponding to overall strength parameter $c_p = 0.25$ and antiparallel flux-orientation suppression factor $\omega = 0.5$. The optional non-linear scaling exponent r (**ClosePacking:expNSP**, with default value 0.5) is a leftover from the original thermal-model implementation [44]. It is not extensively examined here but has been left in the implementation for generality and for consistency with [44].

Closepacking effects on flavour. Given an effective string tension, the next question then becomes how does this effect flavour probabilities used in string fragmentation by PYTHIA. The main set of parameters that will be impacted by a modified string tension are **StringFlav:ProbStoUD** ($P(s : u/d)$, the probability of a strange quark relative to up/down), **StringFlav:ProbQQtoQ** ($P(qq : q)$, the probability of a diquark relative to quark), **StringFlav:ProbSQtoQQ** ($P(sq : qq)$, additional suppression of strange flavour production for diquarks), and **StringFlav:ProbQQ1toQQ0** ($P(qq_1 : qq_0)$, the probability of a spin-1 diquark relative to a spin-0 diquark, additional to the factor 3 enhancement of spin-1 production from state counting).³

³In PYTHIA, up and down quarks are treated on the same footing i.e., with equal probabilities.

Let us first consider how the strangeness probability, $P(s : u/d)$, scales according to the string tension. Given the Gaussian form of the Schwinger mechanism as per eq. (2.1), the strange-to-up/down quark production probability takes the form,

$$P(s : u/d) = \frac{P(m_s^2)}{P(m_{u/d}^2)} = \frac{\exp\left(-\frac{\pi m_s^2}{\kappa_0}\right)}{\exp\left(-\frac{\pi m_{u/d}^2}{\kappa_0}\right)} = \exp\left(-\frac{\pi(m_s^2 - m_{u/d}^2)}{\kappa_0}\right). \quad (2.8)$$

Similarly, given an effective string tension, the modified probability $P'(s : u/d)$ is then given by,

$$P'(s : u/d) = \exp\left(-\frac{\pi(m_s^2 - m_{u/d}^2)}{\kappa_{\text{eff}}}\right). \quad (2.9)$$

This allows us to rewrite the probability $P'(s : u/d)$ in terms of $P(s : u/d)$ and the ratio of string tensions $\kappa_0/\kappa_{\text{eff}}$,

$$P'(s : u/d) = \exp\left(-\frac{\pi(m_s^2 - m_{u/d}^2)}{\kappa_0} \frac{\kappa_0}{\kappa_{\text{eff}}}\right) = P(s : u/d)^{\kappa_0/\kappa_{\text{eff}}}. \quad (2.10)$$

Importantly, the effective string tension κ_{eff} as given in eq. (2.7) is proportional to the vacuum string tension κ_0 , meaning that the ratio $\kappa_0/\kappa_{\text{eff}}$ is independent of the vacuum string tension and hence this scaling of the strangeness probability is parameterized insensitive to both the uncertainty in the string tension value and the quark masses. The implications of eq. (2.10) and the modified string tensions is evident in figure 8, which shows the fraction of strange hadrons given various multiplet representations. Interestingly, the 10-plet ($p = 3, q = 0$) and 27-plet ($p = 2, q = 2$) representations result in the same strange hadron fractions. Though being different colour multiplets with differing colour factors ($C_{10} = 4.5C_F$ and $C_{27} = 6C_F$), the per string contribution to the effective tension for both configurations is 1.5, and thus the same level of strangeness enhancement is expected as evident in figure 8.

The probabilities `StringFlav:ProbSQtoQQ` and `StringFlav:ProbQQ1toQQ0` also are expected to be modified in the same manner as per eq. (2.10). The strength of the closepacking mechanism, c_p in eq. (2.7), for the aforementioned flavour probabilities is controlled by parameter `ClosePacking:enhanceStrange`.

The modification of the diquark probability `StringFlav:ProbQQtoQ` is less straightforward however. There remains ambiguity whether the probability should be scaled assuming the diquark is treated as a constituent formed via the Schwinger mechanism, or whether it should be modelled via the popcorn mechanism. In section 2.4 we will explore further the physics of the popcorn mechanism given dense string environments, however here let us first consider how the probability would be scaled assuming Schwinger-type diquark production.

The diquark probability $P(qq : q)$ can be parameterized in terms of probabilities $\mathcal{P}_{ud0}$, $\mathcal{P}_u$, and a multiplicative factor α which is a function of the probabilities $P(s : u/d)$, $P(sq : qq)$, and $P(qq_1 : qq_0)$. The full functional form is seen in appendix A.

$$P(qq : q) = \alpha \frac{\mathcal{P}_{ud0}}{\mathcal{P}_u}. \quad (2.11)$$

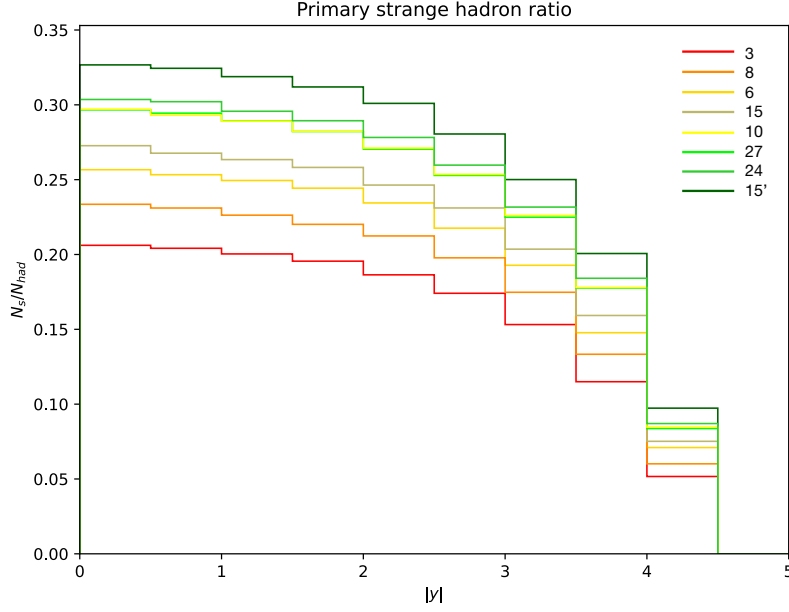


Figure 8. Strange hadron probabilities with respect to absolute rapidity for different string multiplet configurations. These multiplet configurations are constructed using parallel $u\bar{u}$ strings along the beam axis, with endpoint energies of 20 GeV.

Assuming the term $\mathcal{P}_{ud0}/\mathcal{P}_u$ scales with the a power of $\kappa_0/\kappa_{\text{eff}}$, the modified diquark probability $P'(qq : q)$ can be written as

$$P'(qq : q) = \tilde{\alpha} \left(\frac{P(qq : q)}{\alpha} \right)^{\frac{\kappa_0}{\kappa_{\text{eff}}}}, \quad (2.12)$$

with $\tilde{\alpha}$ having the same functional form as α whilst taking into account the effective string tension. In principle, then one can scale $P(qq : q)$ using the same $\kappa_0/\kappa_{\text{eff}}$ used to modify the aforementioned probabilities such as $P(s : u/d)$. However given the additional ambiguity afforded to diquark creation — for example the uncertainty between the Schwinger-type production and popcorn mechanism — we allow the closepacking strength on diquark production — i.e., c_P in eq. (2.7) — to not match that used by the other flavour probabilities. Thus we introduce additional parameter `ClosePacking:enhanceDiquark`, which sets c_P used in calculating κ_{eff} for modifying `StringFlav:ProbQQtoQ`. Note that even with `ClosePacking:enhanceDiquark` set to zero, due to $\tilde{\alpha}$ having dependence on the other flavour probabilities, the diquark probability will still be altered by closepacking. Thus to further capture the ambiguity of diquark production modification, we introduce parameter `ClosePacking:doEnhanceDiquark`, which allows for any modification of the diquark probability due to closepacking to be switched off entirely.

Closepacking effects on $\mathbf{p}_\perp$. Due to the Gaussian form of the Schwinger mechanism with respect to transverse mass, the transverse momentum spectra for string breaks should also be effected by string tension changes. In PYTHIA, the width of the $p_\perp$ spectrum of string breaks is given by σ^2 , with σ set by parameter `StringPT:sigma`. In terms of the string tension, the

average transverse momentum can thus be written as

$$\sigma^2 = \langle p_\perp^2 \rangle = \frac{\pi}{\kappa_0} \int_0^\infty p_\perp^2 \exp\left(\frac{-\pi p_\perp^2}{\kappa_0}\right) dp_\perp^2 = \frac{\kappa_0}{\pi}. \quad (2.13)$$

Given this simple dependence on the string tension, the modified width can be written as per eq. (2.14) in terms of the ratio of string tensions, effectively scaling `StringPT:sigma` by a multiplicative factor of $\sqrt{\kappa_{\text{eff}}/\kappa_0}$.

$$\sigma'^2 = \frac{\kappa_{\text{eff}}}{\pi} = \frac{\kappa_0}{\pi} \frac{\kappa_{\text{eff}}}{\kappa_0} = \sigma^2 \frac{\kappa_{\text{eff}}}{\kappa_0}. \quad (2.14)$$

To calculate the effective tension used for string-break $p_\perp$ -distribution modifications, `ClosePacking:enhancePT` is used to set the strength parameter c_P used in eq. (2.7).

2.3 Strange junctions

The apparent surplus of proton production in PYTHIA given the overprediction of p/π ratio in conjunction with the enhancement of multi-strange baryons seen by ALICE [32] is suggestive of strangeness enhancement effects possibly being focused in the baryon sector. Evidently though, the strangeness enhancement effects are not exclusively in the baryon sector as can be seen by the strange meson ratio increases (e.g. K_s/π and ϕ/π ratio increases with multiplicity), however the level of increase in the multi-strange baryon sector certainly brings into question whether there are different considerations for strange baryon production. In this section, we explore the possibility of strange baryon enhancement being sourced from junction-baryon formation specifically.

One approach is to exploit the ambiguity around what the properties of the colour field near an SU(3) junction are, a question which has recently been explored, e.g., in [59]. The simple sketch in figure 9 illustrates that one may question whether the properties of the confinement field around a junction are identical to those in a simple dipole string in vacuum. Within the closepacking paradigm, one may suppose that the large observed enhancements of strange baryons in particular could in some way be related to an enhanced energy density per unit length around a junction, which would primarily affect string breaks “near” the junction. This in turn would focus strangeness enhancement around junctions, and hence show up primarily in the baryon sector.

Of course in order to create a $s\bar{s}$ pair from a string break next to a junction, if the junction baryon contains the s quark then the next adjacent hadron must contain the $\bar{s}$ antiquark, meaning this mechanism would enhance not only junction-baryon strangeness but also meson strangeness “near” junctions. Nonetheless it does ensure a focus on strangeness in the baryon sector, and the relative scaling of the production of different strange hadrons species (e.g., with respect to variables such as charged multiplicity) will be varied comparative to other strangeness enhancement models such as closepacking.

We will refer to this ambiguity of the junction field strength as “strange junctions” throughout the remainder of this text. In the implementation of strange junctions, we allow the strangeness enhancement to only be effective for the first string break next to the junction itself. In PYTHIA, this can be scaled by the parameter $J_s \in [0, 1]$ (`StringFragmentation:enhanceStrangeJunction`). $J_s = 1$ corresponds to maximal

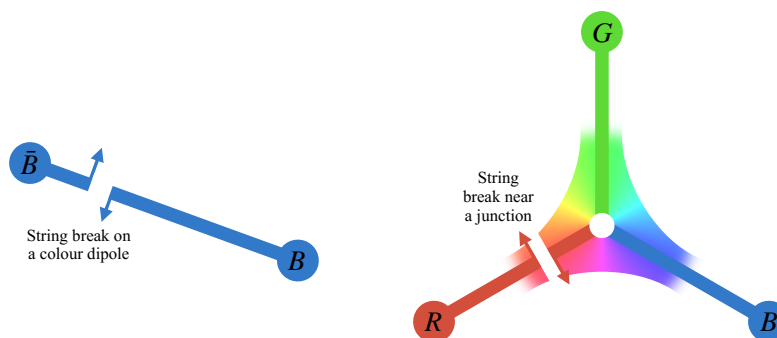


Figure 9. Here we show side-by-side a dipole and junction string to demonstrate the motivation for the strange-junction model. On the left is a simple colour dipole string in vacuum with a single string break. On the right is a junction string topology, with a string break near the junction. The ambiguity of the colour field description surrounding the junction itself is pictorially represented by the continuous colour field between junction legs nearby the junction. This of course is only a pictorial example to convey such ambiguity with possible unknown properties close to the junction, and by no means should be taken as an exact representation of what the junction field should look like.

strangeness around junctions, while $J_s = 0$ corresponds to no (additional) strangeness enhancement around junctions. As such, the strangeness probability for breaks next to the junction are modified according to,

$$P'(s : u/d) = P(s : u/d)^{1-J_s}. \quad (2.15)$$

This additional strangeness enhancement has been implemented subsequent to closepacking effects on the strangeness probability.

2.4 Popcorn destructive interference

In this section we explore the expected modification of diquark production in dense string environments given the popcorn mechanism [51, 52]. Contrary to the closepacking assumption of diquark probabilities scaling upwards in higher-density string systems, here we consider a model which drives the diquark production rate in the opposite direction. Building upon the existing popcorn mechanism, we introduce a novel mechanism which instead reduces diquark formation via a combination of physics and colour-algebra arguments.

In the popcorn mechanism as outlined in section 2.1, diquark production is described by successive colour fluctuations on a string that are allowed to form real diquarks given the necessary colour configurations. Given multiple strings in close proximity, suppose these virtual colour fluctuations are allowed to colour-connect to other nearby strings prior to diquark formation occurring. This would mean that a colour fluctuation on one string could allow for the formation of a real quark-antiquark break on another. Effectively this would prevent a portion of diquarks from forming, reducing the overall diquark — and hence baryon — production rate.

To illustrate this notion, let us consider the same example as per figure 2; a red-coloured string with an initial blue-antiblu colour fluctuation. Now consider the same scenario but also with a blue-colour string piece nearby as shown in figure 10(a) - (b). Should the nearby

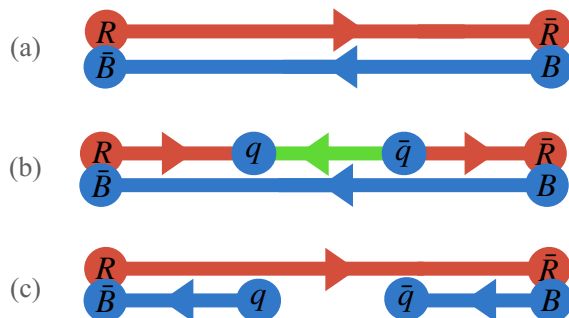


Figure 10. An illustration of the stepwise process of the popcorn mechanism given a red-antired string. (a) the unbroken red-antired string next to a blue-antibblue string with oppositely oriented flux. (b) a blue-antibblue virtual colour fluctuation on the red-antired string, which results in a short segment of “green-coloured” string. (c) the blue-antibblue fluctuation breaks the nearby blue string.

blue string be able to form a colour singlet with the blue-antibblue colour fluctuation formed on the red string, suppose the fluctuated quark pair was allowed to form a real break the nearby blue-colour string with some probability. This processes is illustrated in figure 10(c), and will thence prevent a fraction of diquark formation on the red string.

To implement this so-called popcorn destructive interference effect in PYTHIA, rather than dynamically modelling the colour fluctuations themselves, the net effect is modelled via a reduced diquark production probability that scales with the density of the string environment. To do so, we require some probability that a given virtual colour fluctuation will colour-connect with an overlapping string provided the correct colour configuration. Parameters `ClosePacking:baryonSup` ($\mathcal{P}_q$) and `ClosePacking:parallelBaryonSup` ($\mathcal{P}_p$), give this interference probability for antiparallel and parallel string orientations respectively. The setting of such probabilities for each flux orientation is included as flux sensitivity of this mechanism is reasonably expected. As evident in the scenario outlined in figure 10, in order for the blue-antibblue colour fluctuation to become a real break on the blue string, the favourable orientation is for the blue string to be antiparallel to the red.

Important to note is that in the above thus far we have simply represented colour as red, green, or blue. However more realistically one needs to consider accurately describing SU(3) colour-space probabilities. According to SU(3) colour algebra we have $3 \otimes \bar{3} = 8 \oplus 1$, which effectively gives a 1/9 probability of a given colour-anticolour pair forming a singlet state. This means that for n nearby strings, we expect on average $n/9$ strings to have the correct colour configuration to form a colour singlet with a given virtual colour fluctuation. Per nearby string with the necessary colour configuration, the survival probability of a given fluctuation is $(1 - \mathcal{P})$. As the fluctuation needs to survive each nearby string, with the inclusion of flux sensitivity the modified diquark formation probability is therefore defined by,

$$P'(qq : q) = P(qq : q)(1 - \mathcal{P}_q)^{n_q/9}(1 - \mathcal{P}_p)^{n_p/9}, \quad (2.16)$$

with n_p and n_q being the total number of overlapping parallel and antiparallel strings respectively.

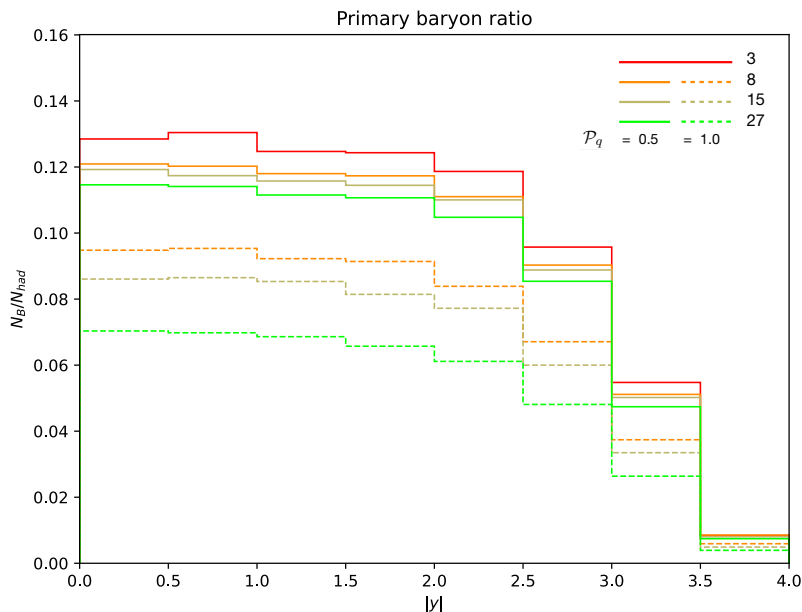


Figure 11. Baryon probabilities with respect to absolute rapidity for various string multiplet configurations, with string configurations constructed using parallel $u\bar{u}$ strings along the beam axis, with endpoint energies of 20 GeV.

Additionally, similarly to closepacking effects being lessened when strings are further apart from one another, one would expect the same notion here with the probability of a fluctuation’s colour-connection to a nearby string reducing with separation. Using the same trial-hadron method as for closepacking, with the same $p_\perp$ suppression form on the connection probabilities $\mathcal{P}_{q/p}$, we get the full form of the modified diquark probability according to popcorn destructive interference,

$$P'(qq : q) = P(qq : q) \left(1 - \frac{\mathcal{P}_q}{1 + \frac{p_{\perp had}}{p_{\perp 0}}} \right)^{n_q/9} \left(1 - \frac{\mathcal{P}_p}{1 + \frac{p_{\perp had}}{p_{\perp 0}}} \right)^{n_p/9}. \quad (2.17)$$

Figure 11 provides proof of concept for the implementation of the popcorn destructive interference effect, clearly demonstrating that the baryon fractions reduce for higher multiplet representations, and more so for larger values of $\mathcal{P}_q$.

3 Constraints and results

In this section, we will refine the parameters used for the models introduced in the preceding text, with the primary aim to fit the ALICE strange hadron-to-pion ratios simultaneously with the typically overpredicted proton-to-pion ratio [32]. We will make comparisons to the existing Monash 2013 tune [11], QCD CR (time dilation mode $m_\tau = 2$) [45], the Rope hadronization model [36, 37], and the closepacking tune from [60] which we label the “Trieste 1” tune, or T1 for short. Below we shall give brief overviews of each of these models and key aspects that shall be pertinent to the following comparison and tuning discussions. We then elaborate on some of the constraints imposed on the aforementioned new models, and outline the methodology used to drive our tuning procedure.

We perform three separate tunes, all using QCD CR ($m_\tau = 2$) as the foundation; one with only the strange junctions model turned on (which we shall label the “Strange junctions” tune), one with only closepacking turned on (“Closepacking (simple)”), and a third tune with closepacking, popcorn destructive interference, and strange junctions all used in combination, which we label the Closepacking Trieste 2 tune (“Closepacking (T2)”).

3.1 Models for comparison

The Monash tune is the default parameter set for PYTHIA 8.3 [3] and uses the standard LSM [1]. It utilizes a simple (so-called “MPI-based”) colour-reconnection model, for which the modelling of colour space remains leading colour (i.e., number of colours $N_C \rightarrow \infty$). Accordingly, this CR model does not lead to the creation of any (new) colour junctions. With the exception of baryon beam remnants [61, 62] (of which implications will not be elaborated on within the scope of this paper), the only source of baryon formation is diquark-antidiquark pair production in string breaks. The fragmentation parameters in the Monash tune are tuned to measurements of identified-hadron yields and spectra in hadronic Z decays at LEP, of which it describes quite well [11]. Nonetheless, the model falls short at describing pp collision data. Of important notice is the constant diquark and strangeness probabilities along a string, which are fixed by parameters `StringFlav:ProbQQtoQ` and `StringFlav:ProbStoUD` respectively. This consequently results in relatively constant distributions of strange hadron-to-pion and baryon-to-meson ratios with respect to both $p_\perp$ and charged multiplicity as would be expected from the baseline LSM. This no longer remains adequate in pp collision environments however, evidenced by many LHC results. A clear example of such discrepancy in the strange sector is given by the strangeness enhancement with respect to charged multiplicity observed by the ALICE collaboration [32, 33], as demonstrated by figure 13. Interestingly, despite the diquark production rate `StringFlav:ProbQQtoQ` being tuned LEP events, the Monash tune overpredicts the proton-to-pion ratio in pp collision environments as seen in figure 13(c).

The next model we compare to in this paper is called the QCD CR model [45]. The key defining feature of this model is the stochastic reintroduction of colour-space ambiguities according to SU(3) algebra, including the formation of so-called junction string topologies [45–47]. Details of the implementation of this beyond-leading-colour model will not be outlined in this paper, however for a simple review on QCD CR see [35]. In particular, we will compare to a variation of the tune called “CR (time-dilation mode $m_\tau = 2$)”, with parameters listed in appendix C.⁴

In e^+e^- collision events, the corrections of including colour-space ambiguities are expected to be approximately $1/N_C^2 \sim 10\%$. In pp collision events however, denser partonic systems allow for more colour space overlaps, thus resulting in enhanced QCD CR effects. Of particular interest is the significance of the inclusion of junction strings as an additional baryon production mechanism. Junction baryons have proven effective in the description of the Λ/K_S ratio [49] in pp collisions, alongside well describing heavy flavour baryon distributions such as the Λ_c/D^0 ratio [63] and the Λ_b asymmetry [64] with respect to transverse momentum.

⁴The parameter values for `StringFlav:probQQ1toQQ0join` used throughout this paper are varied from the initial tuning on QCD CR ($m_\tau = 2$) as per [45]. Instead we return to the default PYTHIA values for `StringFlav:probQQ1toQQ0join` in order to better fit the $\Sigma_c^{0,+,++}/D^0$ and $\Lambda_c^+(\leftarrow \Sigma_c^{0,+,++})/\Lambda_c^+$ ratios (see [35]).

Despite the success of describing the aforementioned baryon distributions, there remain several shortcomings to the model. Given the Monash tune already overpredicts the p/π ratio in pp collisions, the inclusion of junction baryons only further pronounces such an overprediction as demonstrated in figure 13(c). Additionally, QCD CR does not provide any strangeness-enhancement mechanism. CR effects — including junction-baryon formation — are correlated with the density of the partonic environment, of which the charged multiplicity (N_{ch}) is also correlated. This means that the model results in an enhancement of baryon production with respect to charged multiplicity, however such baryon enhancement cannot describe the level of enhancement required for multi-strange baryon-to-meson ratios, nor can the observed enhancement of strange meson-to-pion ratios (e.g. K_S/π and ϕ/π shown in figure 13(a) and figure 13(b) respectively) be explained.

Another model we shall compare to here is Rope hadronization [36, 37]. The Rope hadronization model is designed to produce strangeness enhancement and thus is important for comparison here. In this model, a collection of nearby strings is treated as a coherent system — namely a “rope” — which then fragments with an enhanced string tension correlated to the number of strings. Consequently the model results in enhancement of both strangeness and diquark production with respect to charged multiplicity. Given the default values of the Rope hadronization model, sufficient levels of strangeness enhancement are present in order to fit both the strange meson K_S/π ratio (see figure 13(a)) alongside the multi-strange baryon Ω/π ratio. Nonetheless, there remains a problem with the overprediction of the p/π ratio (see figure 13(c)).

One should also note that the Rope hadronization model takes a similar approach to strangeness enhancement which we have taken in our closepacking model outlined in section 2.2, however closepacking aims to be a somewhat simplified model of the same principle. The key difference in the two approaches lies in the treatment of the surrounding strings. The Rope hadronization model assumes the formation of coherent colour-rope structures, whereas the closepacking model simply treats the surrounding strings as an effective background on a given fragmenting string. This variance in approaches means that whilst the closepacking model remains a fully momentum-space implementation, Rope hadronization requires space-time mappings of string breaks. Such a fully momentum-space fragmentation procedure for closepacking allows for both a more natural compatibility with junction topologies, alongside a reduced execution speed.

The final tune we shall compare to is the earliest closepacking tune obtained in [60], using a version of the closepacking implementation from PYTHIA 8.311. We will label this tune “Closepacking (T1)” throughout the following.

For completeness, parameter specifications for each of the reference models mentioned above are given in appendix C.

3.2 Tuning methodology

Given the key interest of the above collectivity models is to reproduce effects expected in pp collisions, the QCD CR (time-dilation mode $m_\tau = 2$) [45] will be the basis for the proceeding tune. We will focus on describing the ALICE strange hadron-to-pion ratios simultaneously with the proton-to-pion ratio, whilst maintaining consistency with other global pp observables

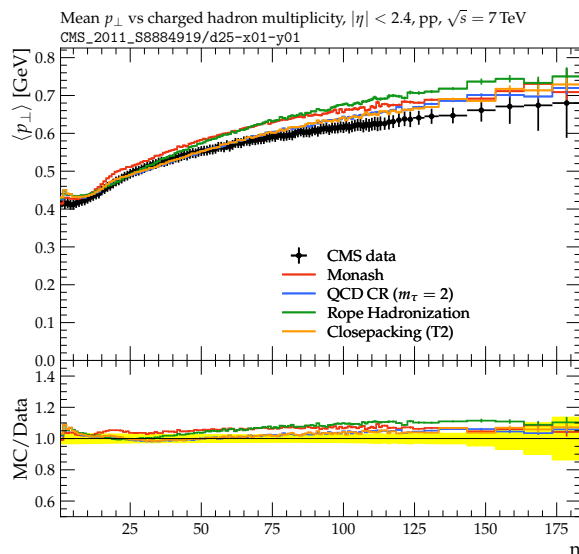


Figure 12. Average $p_{\perp}$ with respect to charged multiplicity for 7 TeV NSD events at CMS [65]. PYTHIA 8.316 simulations for tunes listed in table 6 are run with a lifetime cut of $\tau_{\text{max}} = 10\text{mm}/c$, and no $p_{\perp}$ cuts on final-state particles.

such as $\langle p_{\perp} \rangle$ (see figure 12 and 14) and $dn_{ch}/d\eta$. This means that we require the tuning of the additional new parameters introduced for closepacking modelling, additional to the retuning of MPI and CR parameters.

In the following, we will use Professor [10] to find best fit values for parameters. The workflow with Professor can be divided into four steps; sampling, event generation, interpolation, and χ^2 minimization. The first one consists in *sampling* the parameter space of the P parameters that one wants to tune, defining the corresponding variation ranges $\mathbf{p}_{\text{min}}$ and $\mathbf{p}_{\text{max}}$ (using the same notation of the original paper [10]). By default, sampling is performed uniformly within the P -dimensional hypercube $[\mathbf{p}_{\text{min}}, \mathbf{p}_{\text{max}}]$. At each sampled point, the MC generator is run for the desired Rivet [66] analyses, producing histograms saved as YODA files [67]. Upon conclusion of the simulations, the Professor tool can read all the Rivet histograms corresponding to the different parameter values. It then performs a fit of the MC response $f_b(\mathbf{p})$, specifically for the same bin b of each observable $\mathcal{O}$, using polynomial functions of the desired order. This process, known as *interpolation*, yields a continuous parameterization of the MC response, as a function of the tuning parameters. The tuning results presented in this work were obtained using 4th- and 5th-order polynomials for the interpolation step. The higher the degree of the polynomial and the number of parameters to tune, the higher the number of sampled points required for the interpolation step. For example, in a 4-dimensional parameter space, the minimum number of sampled points required is 35 with 3rd-order polynomials. This number increases to 70 and 126 using 4th and 5th-order polynomials, respectively. Moreover, [10] suggests an oversampling by a factor of 2.

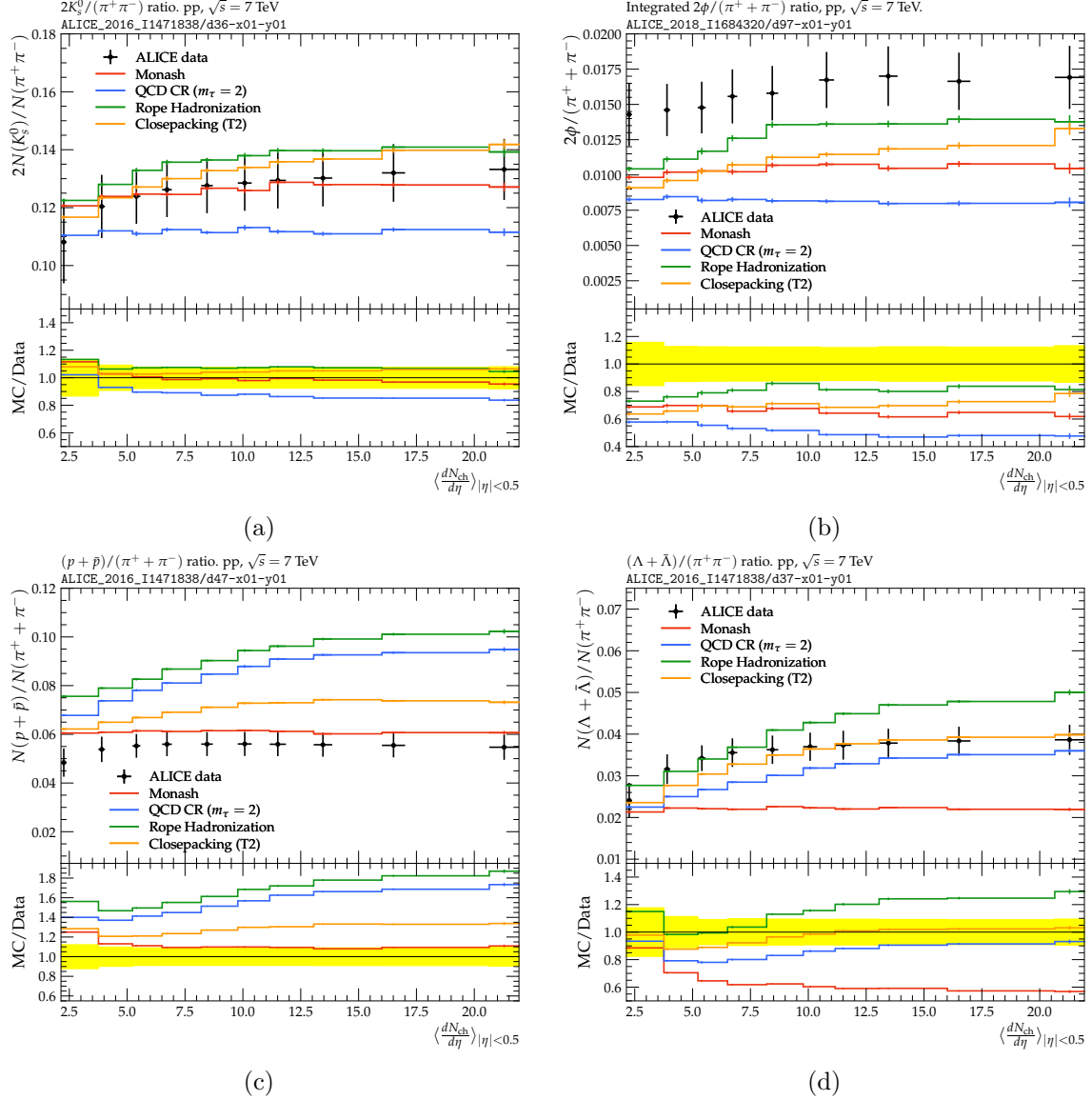


Figure 13. Hadron-to-pion ratios with respect to ALICE midrapidity charged multiplicity classes. PYTHIA 8.316 simulations for parameters in table 6 of 7 TeV INEL>0 events are run with a lifetime cut of $\tau_{\text{max}} = 10\text{mm}/c$, and no $p_{\perp}$ cuts on final-state particles. Top left: $2K_S^0/\pi$ [32]. Top right: $2\phi/\pi$ [33]. Bottom left: p/π [32]. Bottom right: Λ/π [32].

The parameterization of the MC is subsequently used to construct a weighted χ^2 along with the experimental data:

$$\chi^2 = \sum_{\mathcal{O}} w_{\mathcal{O}}^2 \sum_{b \in \mathcal{O}} \frac{(d_b - f_b(\mathbf{p}))^2}{\sigma_{d_b}^2 + \sigma_{f_b}^2}, \quad (3.1)$$

where d_b are the data for the bin b of the observable $\mathcal{O}$, which are automatically retrieved by Professor from Rivet [66]. Variables σ_{d_b} and σ_{f_b} are the errors associated to the experimental measurement and to the MC parameterization, respectively, and $w_{\mathcal{O}}$ is the weight associated to the observable $\mathcal{O}$ as chosen by the user.

To prevent overfitting a flat 5% “theory uncertainty” [11–13, 68] is applied to the χ^2 as follows:

$$\chi_{5\%}^2 = \sum_{\mathcal{O}} w_{\mathcal{O}}^2 \sum_{b \in \mathcal{O}} \frac{(d_b - f_b(\mathbf{p}))^2}{\sigma_{d_b}^2 + \sigma_{f_b}^2 + (0.05 f_b)^2}, \quad (3.2)$$

For the purpose of this tuning, $\chi_{5\%}^2$ is employed. The minimization of this function performed with Professor returns the vector of the best parameter values $\mathbf{p}_{\text{best}}$. Throughout this work YODA 1.9.10, Rivet 3.1.10 [66], Professor 2.4.2 [10] are used. Plots comparing different tunes predictions are obtained using the `rivet-mkhtml-mpl` script with some additions. Moreover, for each plot a goodness of fit is computed and shown for each observable using a similar definition implemented in CERN MCPLOTS [14, 69],

$$\chi_{\text{GoF}}^2 = \frac{1}{N_{\text{bins}}} \sum_{b=1}^{N_{\text{bins}}} \frac{(d_b - \text{MC}_b)^2}{\sigma_{d_b}^2 + \sigma_{\text{MC}_b}^2 + (0.05 \text{MC}_b)^2}, \quad (3.3)$$

where N_{bins} is the number of bins of the histogram, MC_b the value of the MC prediction in bin b and σ_{MC_b} its corresponding uncertainty. A 5% “theory uncertainty” is also added in this case. In table 2, the goodness of fit level is shown for various observables used in the tuning procedure for each of the tunes, using a gradient colour scheme from green to purple/black, with green corresponding to a low χ_{GoF}^2 value, and purple a high χ_{GoF}^2 value.

We note that, since most of the experimental measurements we compare to are dominated by systematics, we expect $\chi_{\text{GoF}}^2 \ll 1$ for a model that goes “straight through” the data, rather than $\chi_{\text{GoF}}^2 \sim 1$ as would be the case if the data were statistically scattered within their uncertainty bands. However, our main focus here will be on which cases exhibit $\chi_{\text{GoF}}^2 \gg 1$. We do not mean to imply that a model that achieves $\chi_{\text{GoF}}^2 \sim 0$ is much better than one that achieves $\chi_{\text{GoF}}^2 \sim 1$ although the χ_{GoF}^2 minimiser will of course select for the lowest possible values.

To constrain the dimensionality of the problem, the tuning procedure has been split into several stages with each targeting a focused subset of observables. After each stage in this tuning procedure, parameters are then either fixed according to the best fit or allowed to remain floating with a more tightly constrained range in the proceeding tuning stage. Each stage of this tuning methodology has been chosen in order to aim to minimize the correlation between parameters across stages, with the target areas as follows (listed in order of approach); MPI and CR levels, baryons and strangeness, and $p_{\perp}$ dependence.

Preceding the explanation of the tuning scheme, below we outline which parameters are assumed as fixed in our tune, and the justification for such settings. Firstly, in order to remain consistent with strange/diquark production rates expected in e^+e^- collisions, flavour probabilities such as `StringFlav:probStoUD` and `StringFlav:probQQtoQ` will remain fixed according to the QCD CR ($m_{\tau} = 2$) tune. In principle, closepacking may effect a subset of e^+e^- events that involve nearby jets, however we expect these effects to be minimal. Regardless, the existing model for both closepacking and popcorn destructive interference is designed only for application to hadronic collisions given the beam axis dependence which is based on the assumption that the bulk of strings are oriented along the beam axis. This

therefore means the models are not yet useful for describing e^+e^- collisions. Generalization of the model applications to arbitrary collision environments remains work for a future study.

We also acknowledge that the newly introduced parameters that dictate flux sensitivity of the models — `ClosePacking:parallelBaryonSup` and `ClosePacking:fluxRatio` — cannot be finely tuned given the lack of flux-sensitive experimental observables available. In order to tune such flux sensitivity, one would need to be able to differentiate between multiplet configurations in cleaner collision environments, such as e^+e^- collisions, and then compare hadronic distributions sourced from the fragmentation of e.g. an octet vs a sextet configuration. Given we do not have such experimental observables at hand, we instead make reasonable default assumptions for the values of the flux-sensitive parameters. For the popcorn destructive interference modelling, we assume `ClosePacking:parallelBaryonSup` ($\mathcal{P}_p$ per eq. (2.17)) is always set to zero, so that the contribution of the interference effect is isolated to antiparallel string configurations. This is physically sensible as antiparallel strings are the favourable orientation for the mechanism to occur and would thus be expected to be the dominating contribution. For tension modifications via closepacking, `ClosePacking:fluxRatio` (ω per eq. (2.7)) is fixed at 0.5, which is aligned to the strength of flux sensitivity expected by Casimir scaling.

Additionally, we set `ClosePacking:doEnhanceDiquark = off`. This prevents the string-tension modifications due to closepacking from scaling the diquark probability upwards in a manner assuming Schwinger-type production of diquarks per eq. (2.12). Instead, as we wish to simultaneously implement closepacking alongside the popcorn destructive interference mechanism introduced in section 2.4, we must then assume popcorn-type diquark formation which one would not necessarily expect to scale in the same way as eq. (2.12). How exactly popcorn diquark formation probabilities should scale with an increased string tension remains somewhat ambiguous. Regardless, if one were to allow the diquark probability to increase due to enhanced tensions, this effect would then be directly counteracted by popcorn destructive interference, and thus it would not make sense to allow both scalings to occur in conjunction. As the popcorn destructive interference has not been dynamically implemented, but rather the effectively implemented via a suppression factor on the diquark probability controlled by `ClosePacking:baryonSup`, we allow this parameter to holistically capture the effect of surrounding strings on the diquark production rate — incorporating the combination of enhanced string tensions from closepacking and popcorn destructive interference effects. This of course assumes a net decrease in the diquark production rates. Should one wish to assume a net enhancement, one should set `ClosePacking:baryonSup = 0`, and use `ClosePacking:doEnhanceDiquark = on` whilst allowing `ClosePacking:enhanceDiquark` to control the effective diquark enhancement strength. In the following however, we will assume the diquark probability decreases with the density of the string system.

We also fix the closepacking regularization parameter $p_{\perp 0}$ (`ClosePacking:PT0`) to a value of 1.75 GeV. This is partially as we do not possess experimental results for observables that would be particularly sensitive to this parameter. Ideally for tuning of this parameter, one would want two (anti)parallel strings with varying separations. For the $p_{\perp}$ observables we do have, sensitivity to the exact value of this parameter is not large. Given the hadronic $p_{\perp}$ distributions are not particularly well describe in general, we do not fine-tune this parameter.

Expecting it to sit somewhere between 1–3 GeV, we fix it 1.75 GeV in this work, implying that the additional tension modification would halve for a $p_{\perp\text{Had}}$ of 1.75 GeV.

The final parameter which we fix is `ColourReconnection:mPseudo` (m_{pseudo}). This is a newly introduced parameter as of PYTHIA 8.311. Formerly, both of the parameters `ColourReconnection:m0` and `ColourReconnection:mPseudo` were set by the same parameter labelled `ColourReconnection:m0`. These have been since factorized given their physical meanings should be treated as different. `ColourReconnection:m0` is now only used as a regularization parameter in the string length measure, and in this tuning scheme is fixed to 0.5. The parameter `ColourReconnection:mPseudo` imposes a cut on the invariant mass of pseudo-particles that are not colour reconnected, and is primarily constrained by the $\langle p_{\perp} \rangle$ vs n_{ch} distribution. In the following tuning methodology, m_{pseudo} was fixed to the set of values 0.2, 0.4, and 0.6 in tuning steps 1 and 2. The best of these values, $m_{\text{pseudo}} = 0.4$, was then selected and fixed throughout the remainder of tunes. This parameter was not allowed to float throughout the entirety of the tuning as we do not have observables capable of splitting the degeneracy between m_{pseudo} and parameter `ColourReconnection:timeDilationPar`.

With the above assumptions in place, we can now proceed with tuning the remaining parameters from the newly introduced models. Below we will outline each stage of the tuning scheme, including which parameters are allowed to float, which are fixed, and which observables each parameter is primarily constrained by.

Step 1: MPI and CR levels. In this step, we aim to more tightly constrain parameters `MultipartonInteractions:pT0Ref` and `ColourReconnection:timeDilationPar`. A description of each of these parameters and observables to be constrained to is given below:

- `ColourReconnection:timeDilationPar`: controls the strength of CR by suppressing high-boost reconnections, constrained largely by $\langle p_{\perp} \rangle$ vs n_{ch} .
- `MultipartonInteractions:pT0Ref`: regularization scale controlling the amount of low $p_{\perp}$ MPIs, strongly correlated to the total multiplicity and thus allowing tuning to $dn_{ch}/d\eta$.

Though the primary aim is to tune closepacking effects to strange hadron and baryon production rates, one must also revisit more global observables given these modified strange/di-quark production rates will impact MPI and CR-sensitive distributions such as total hadronic multiplicities and average transverse hadronic momenta. Thus a retuning of the relevant MPI and CR parameters with closepacking effects switched on is required. Of course as we have yet to tune the closepacking parameters themselves, in this step of the tuning procedure we instead fix these parameters to reasonable guesses. Doing so should yield relatively stable results in proceeding tuning steps despite $\langle p_{\perp} \rangle$ and $dn_{ch}/d\eta$ distributions having some dependence on closepacking effects. The starting parameters were chosen as follows:

- `ColourReconnection:junctionCorrection` = 1.5,
- `ClosePacking:enhanceStrange` = 0.010,
- `ClosePacking:enhancePT` = 0.040,
- `StringFragmentation:enhanceStrangeJunction` = 0.5,
- `ClosePacking:baryonSup` = 0.8.

These starting values were chosen as reasonable guesses given previously conducted rough tuning attempts.

Step 2: baryons and strangeness. In the second step of tuning, we aim to tune the strange hadron and baryon ratios such as K_S/π , Λ/π , p/π , etc. (see appendix F for full list). In particular, we ensure to have a balance of both strange baryon-to-pion ratios alongside strange meson-to-pion ratios. This ensures we can tune the difference in effects of closepacking via `ClosePacking:enhanceStrange` alongside strange junctions via `StringFragmentation:enhanceStrangeJunction`. Of particular note is the ratio ϕ/π . As the ϕ meson is a double strange meson, the strange junctions mechanism will contribute lesser to the ϕ meson enhancement comparatively to closepacking. This is as strange junctions only contribute to strange quark enhancement for a single quark in the meson formed next to the junction baryon. Contrastingly, closepacking will contribute strangeness enhancement to both quarks equally for ϕ meson production. The parameters `ColourReconnection:timeDilationPar` and `MultipartonInteractions:pT0Ref` are fixed from the previous stage of tuning. The parameters tuned in this step alongside the observables they are sensitive to are as follows:

- `ColourReconnection:junctionCorrection`: modifies how easily junction topologies form via a correction to the string length measure used in colour reconnections, and hence alters the baryon production rates.
- `ClosePacking:baryonSup`: popcorn destructive interference strength parameter, $\mathcal{P}_q$ as per eq. (2.17), sensitive to baryon production rates. Particularly will be tuned to the p/π ratio, which is a non-strange baryon-to-meson ratio that the LEP-tuned diquark production rates overpredicts.
- `ClosePacking:enhanceStrange`: closepacking strength parameter c_P defined in eq. (2.7), used to calculate the effective tension for modifications to the strangeness production rate, $P(s : u/d)$, per eq. (2.10). This is sensitive to all strange hadron ratios.
- `StringFragmentation:enhanceStrangeJunction`: strength parameter for strangeness enhancement around a junction, J_s , as per eq. (2.15). This parameter will be particularly sensitive to multi-strange baryon ratios e.g. Ω/π .

Step 3: $p_\perp$ dependence. Here we aim to constrain the $p_\perp$ dependence of closepacking `ClosePacking:enhancePT`, alongside retuning the CR parameters `MultipartonInteractions:pT0Ref` and `ColourReconnection:timeDilationPar`. The floating parameters from step 2 of the tuning procedure are fixed.

Description of the final parameter to be tuned is below:

- `ClosePacking:enhancePT`: closepacking strength parameter c_P defined in eq. (2.7), used to calculate the effective tension for modifications to the width of the $p_\perp$ distribution of string breaks, σ , per eq. (2.14).

After step 3 of the tuning procedure, steps 2 and 3 can be iterated until the parameter values converge, or when no further significant improvement in the χ^2 (and the overall description of the data) is observed.

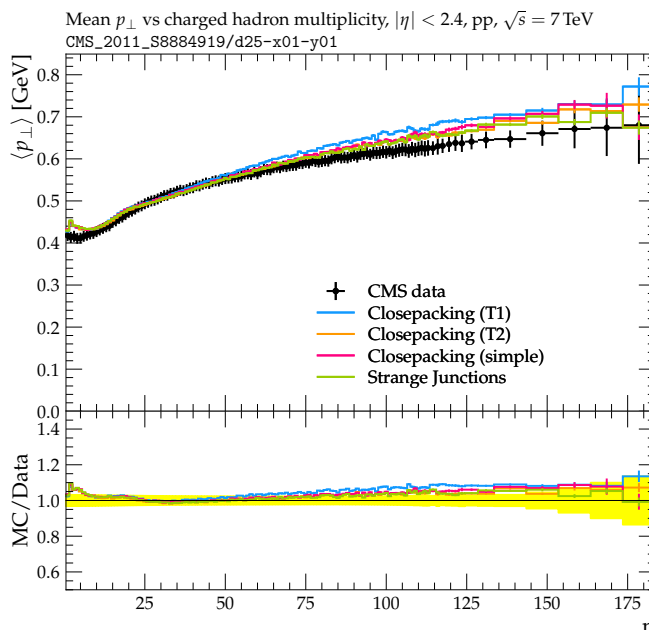


Figure 14. Average $p_{\perp}$ with respect to charged multiplicity for 7 TeV NSD events at CMS [65]. PYTHIA 8.316 simulations for tunes listed in table 1 are run with a lifetime cut of $\tau_{\text{max}} = 10\text{mm}/c$, and no $p_{\perp}$ cuts on final-state particles.

3.3 Tuning outcomes

As stated above, we ran the above tuning methodology⁵ on three different configurations — Strange junctions only (“Strange junctions”), closepacking only (“Closepacking (simple)”), and a combination of closepacking, popcorn destructive interference, and strange junctions (“Closepacking Trieste 2 (T2)”). The parameter values for each of these tunes can be seen in table 1.

To allow for comparison of each models performance, table 2 and table 3 show the goodness of fit values (χ^2_{GoF} per the definition in eq. (3.2)) for each of the model tunes for various observables used in the tuning process. As one can see, the χ^2_{GoF} values show that the Monash and QCD CR tunes fail to describe the majority of strange hadron ratios, and though the Rope hadronization model well describes the strange hadron-to-meson ratios, the p/π ratio is overpredicted. The remainder of the χ^2_{GoF} values also reflect observations pointed out in section 3.1.

Additional to the goodness of fit tables, figure 15 shows various hadron-to-pion ratios given the tune parameters in table 1 alongside Closepacking (T1), with further plots shown in appendix D. The Closepacking (simple) tune performs well for most strange hadron-to-pion ratios, however underpredicts the triple-strange baryon-to-pion ratio Ω/π whilst overpredicting the K_S/π ratio (seen in figure 15(a)). This is somewhat expected given the strangeness enhancement is being driven equally across the mesonic and baryonic sectors, and thus to

⁵Throughout the tuning procedure, there were various bug fixes and improvements were made in the closepacking implementation in PYTHIA. The versions of PYTHIA, the weight configurations for the observables, and the polynomial order used for interpolation for each step in the tuning procedure can be found in appendix F.

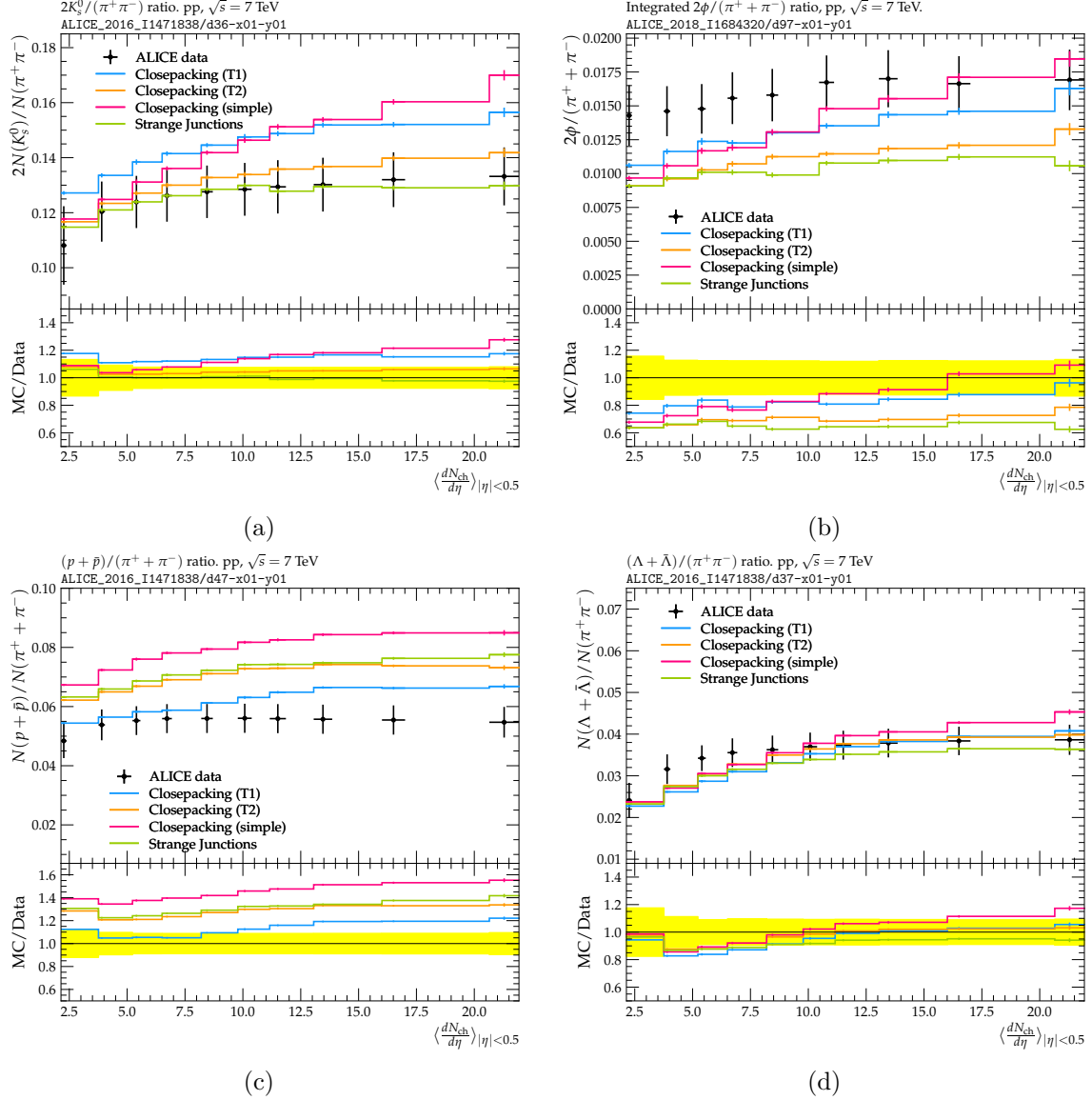


Figure 15. Hadron-to-pion ratios with respect to ALICE midrapidity charged multiplicity classes. PYTHIA 8.316 simulations for parameters in table 1 of 7 TeV INEL>0 events are run with a lifetime cut of $\tau_{\text{max}} = 10\text{mm}/c$, and no $p_{\perp}$ cuts on final-state particles. Top left: $2K_S^0/\pi$ [32]. Top right: $2\phi/\pi$ [33]. Bottom left: p/π [32]. Bottom right: Λ/π [32].

fit the multi-strange baryons, the single-strange mesons are overpredicted. Contrastingly, the Strange Junctions tune performs particularly poorly at describing the double-strange mesonic ϕ/π ratio (see figure 15(b)). This is as the strange junctions mechanism contributes strangeness enhancement around junctions only, and thus applies strangeness enhancement to only a single quark in the meson adjacent to the junction. Therefore the enhancement to the production of double-strange ϕ meson is minimal. Interestingly, the two Trieste tunes (T1 and T2) describe the observables in a contrasting manner despite implementing the same

	Closepacking		Strange
	Trieste 2	Simple	Junctions
MultiPartonInteractions:pt0Ref	= 2.22	= 2.26	= 2.23
ColourReconnection:mpseudo	= 0.4	= 0.4	= 0.4
ColourReconnection:m0	= 0.5	= 0.5	= 0.5
ColourReconnection:junctionCorrection	= 1.67	= 1.68	= 0.96
ColourReconnection:timeDilationPar	= 0.17	= 0.16	= 0.19
ClosePacking:doClosePacking	= on	= on	= off
ClosePacking:PT0	= 1.75	= 1.75	—
ClosePacking:enhanceStrange	= 0.010	= 0.038	—
ClosePacking:enhancePT	= 0.005	= 0.013	—
ClosePacking:doEnhanceDiquark	= off	= off	= off
StringFragmentation:doStrangeJunctions	= on	= off	= on
StringFragmentation:enhanceStrangeJunction	= 0.57	—	= 0.70
ClosePacking:baryonSup	= 0.54	= 0.00	—
ClosePacking:parallelBaryonSup	= 0.00	= 0.00	—

Table 1. Parameter values for tunes using the methodology outlined in section 3.2. All other PYTHIA parameters for each of these tunes are set according to QCD CR ($m_\tau = 2$) [45] with the `StringFlav:probQQ1toQQ0join` values reverted back to the default Monash tune values. Parameters for the other models investigated in this work are given in the appendix, table 6.

	$K_S^0/\pi^\pm$	$\Lambda/\pi^\pm$	$\Xi^\pm/\pi^\pm$	$\Omega^\pm/\pi^\pm$	Λ/K_S^0	$(p + \bar{p})/\pi^\pm$
Monash	0.12	14.18	20.11	31.83	12.32	1.17
QCD CR ($m_\tau = 2$)	1.91	1.98	12.22	18.92	0.58	20.97
Rope Hadronization	0.58	2.40	0.70	0.92	1.11	28.42
Closepacking (T1)	2.14	0.75	0.32	1.98	2.58	1.65
Closepacking (T2)	0.24	0.31	0.11	0.94	0.50	6.37
Closepacking (simple)	2.41	0.76	0.80	2.08	1.07	14.67
Strange Junctions	0.04	0.64	0.15	0.54	0.57	7.72

Table 2. Goodness of fit values, χ_{GoF}^2 , given tunes for various observables used in the tuning procedure. The colour scheme gradients (see table 7) range from green to purple/black, with green corresponding to low χ_{GoF}^2 and purple/black to high χ_{GoF}^2 values.

strangeness enhancement models. The T1 tune describes the p/π and ϕ/π ratios better than the T2 one, but does less well for other ratios such as K_S^0/π and Λ/K_S^0 .

	$\frac{K^++K^-}{\pi^++\pi^-}$	$\frac{K^{*0}+\bar{K}^{*0}}{\pi^++\pi^-}$	$\frac{2\phi}{\pi^++\pi^-}$	$\frac{dN_{ch}}{d\eta}$ vs η	$\langle p_\perp \rangle$ vs n_{ch}
Monash	0.12	0.74	6.76	0.02	0.95
QCD CR ($m_\tau = 2$)	1.67	0.19	13.98	0.24	0.24
Rope Hadronization	0.53	0.89	2.19	0.05	1.19
Closepacking (T1)	1.80	1.63	1.73	0.05	0.66
Closepacking (T2)	0.19	0.65	5.25	0.02	0.24
Closepacking (simple)	1.89	2.58	1.97	0.11	0.31
Strange Junctions	0.05	0.33	7.09	0.02	0.23

Table 3. Goodness of fit values, χ^2_{GoF} , given tunes for various observables used in the tuning procedure. The colour scheme gradients (see table 7) range from green to purple/black, with green corresponding to low χ^2_{GoF} and purple/black to high χ^2_{GoF} values.

The stability and robustness of the new tunes has been tested using a subsampling procedure. This subsampling procedure randomly selects about two thirds of the simulations and performs a tuning procedure on this subset which can then be compared to the tune including all simulations. Results were shown to be stable for most parameters, demonstrating that the tuning is well-constrained and independent of the specific choice of sampled points. Of note is the baryon parameters (`ColourReconnection:junctionCorrection`, `ClosePacking:baryonSup`, and `StringFragmentation:enhanceStrangeJunction`), which had slightly larger dispersions in tuned values between various random subsets. This however is somewhat expected given the level of uncertainties on the baryon data, alongside the lack of available experimental observables which are able to distinguish the junction and diquark baryon sources. Overall however, the results indicate that while some parameters related to baryon production have a broader stability envelope, the overall parameter set is robust and represents a well-defined global minimum.

3.4 Heavy flavour

A full-fledged tuning including heavy-flavour hadron production is not carried out in this work, however in the following we will briefly touch on the effects of the aforementioned strangeness enhancement mechanisms on heavy flavour hadrons and demonstrate how well the tunes perform on key heavy-flavour ratios. In particular, we will focus our attention to examining the large underprediction of the Ξ_c ratios in pp collision environments (see figure 17), particularly at low- $p_\perp$ values.

Though the inclusion of junction baryons has proven to be particularly useful in describing low- $p_\perp$ heavy-flavour baryon production, such as the Λ_c^+/D^0 ratio [63], Ξ_c ratios remain described very poorly. A key example is the 80% underprediction of the Ξ_c^0/D^0 ratio [47, 70]. Of note is the strangeness content of the Ξ_c baryon, which contains both a charm and strange quark in contrast to the Λ_c and D^0 hadrons which have no strangeness content. The vast underprediction of these ratios suggests either an underestimate of Ξ_c baryon production, or the simultaneous overestimation of both the Λ_c and D^0 hadrons. Given charm production is entirely sourced from hard processes and supposing perturbative charm modelling is sound, this leaves the primary assumed source of error to be strangeness

	$\frac{\Lambda_c^+}{D^0}$	$\frac{\Sigma_c}{D^0}$	$\frac{\Lambda_c^+(\leftarrow\Sigma_c)}{\Lambda_c^+}$	$\frac{\Xi_c^0}{D^0}$	$\frac{\Xi_c^+}{D^0}$	$\frac{\Xi_c^0}{\Lambda_c^+}$	$\frac{\Xi_c^{0,+}}{\Sigma_c}$
Monash	41.53	19.83	7.72	8.67	3.14	3.19	0.40
QCD CR ($m_\tau = 2$)	1.69	0.50	0.64	6.06	2.19	5.43	4.00
Rope Hadronization	1.46	0.16	0.42	5.40	2.20	5.11	3.86
Closepacking (T1)	1.66	0.97	0.13	4.54	2.07	4.01	2.67
Closepacking (T2)	1.75	0.54	0.16	5.18	2.22	4.97	3.65
Closepacking (simple)	2.67	0.70	0.40	5.31	2.84	4.36	3.72
Strange Junctions	11.26	6.11	0.79	6.79	3.05	3.87	2.58
Strange Junctions (max.)	12.21	3.67	0.36	6.15	2.60	3.35	2.49

Table 4. Goodness of fit values, χ_{GoF}^2 , of various tunes for heavy-flavour prompt charm observables. The colour scheme gradients (see table 7) range from green to purple/black, with green corresponding to low χ_{GoF}^2 and purple/black to high χ_{GoF}^2 values.

production nearby charm quarks. With approximately 70% of prompt Ξ_c baryons with QCD CR being sourced from junctions — and indeed the majority of any heavy flavour baryons — the study of strange quark production next to junctions themselves become particularly important. Thus in the following, we not only compare to the various Closepacking tunes and Strange Junctions, but compare to an additional set of parameters called “Strange Junctions (max.)” which is simply the equivalent of the Strange Junctions tune however with `StringFragmentation:enhanceStrangeJunction` parameter set to unity. This has been done to test the full extent of the strange junction model on strange charm baryon production specifically, and effectively means that the production of a strange quark is considered equal to that of an up/down quark directly next to a junction. As one can see in figure 17 which is indeed reflected in the χ_{GoF}^2 values in table 4, despite these various strangeness enhancement mechanisms in place, the Ξ_c ratios remain poorly described. Even with this strange junction maximization, the Ξ_c/D^0 and Ξ_c^0/Λ_c^+ ratios are severely underpredicted.

To fully deduce how to well describe these ratios, a dedicated study into the production of charm hadrons would be required, including examination into the string configurations they form and the environment they sit in (i.e., the number of nearby strings and therefore closepacking strength). Such a detailed examination has been left to a future study. One should also reexamine the effect of strangeness enhancement mechanisms on the Λ_b/B^0 ratio [71], which appears to be significantly overpredicted by the QCD CR model [47].

4 Conclusions and outlook

In this study, we have revised and extended the so-called string closepacking model [44] for collective hadronization effects in the PYTHIA 8 Monte Carlo event generator [3]. We have explored the properties of the new model in the context of fitting LHC data, and contrasted it to a set of existing physics modelling options and tunes.

The original closepacking model was introduced in the context of a model for thermal string breaks [44]. In this work, we have:

- extended the closepacking model to apply not only to thermal string breaks but also to conventional Schwinger-type ones;

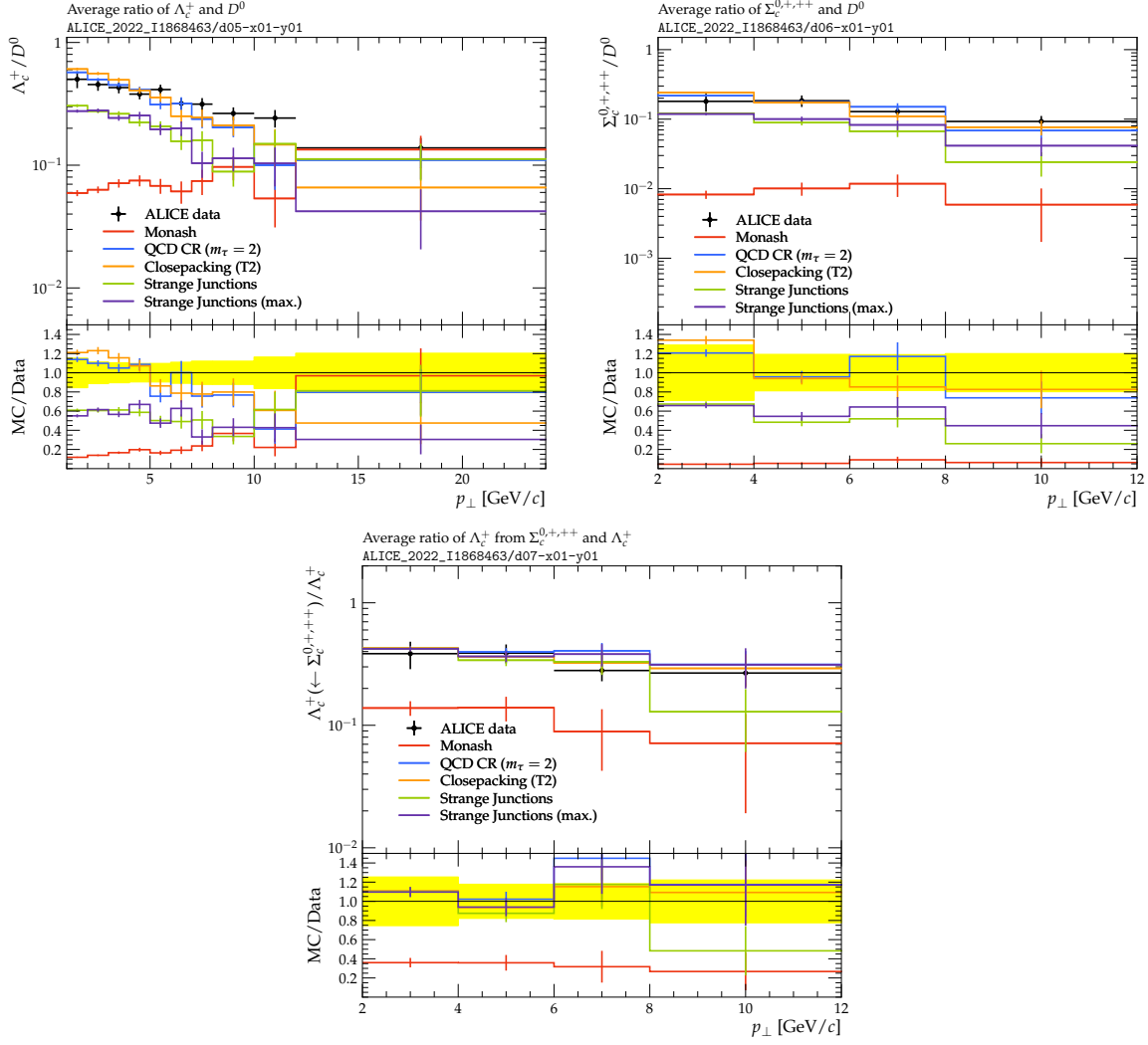


Figure 16. Non-strange charm hadronic ratios, measured by ALICE [63] for $s = 13$ TeV pp collision inelastic events. Measurements are conducted at midrapidity, $|y| < 0.5$, for prompt charm hadrons. Top left: Λ_c^+/D^0 . Top right: $\Sigma_c^{0,+,++}/D^0$. Bottom: $\Lambda_c^+(\leftarrow \Sigma_c^{0,+,++})/\Lambda_c^+$.

- introduced sensitivity to the flux direction, so that not only the total number of strings but also their (relative) flux orientations affect the scaling of the effective string tension, adopting a baseline expectation in line with Casimir scaling;
- proposed a mechanism — popcorn destructive interference — that may reduce diquark production in regular string breaks, based on an argument that nearby colour fields may affect virtual colour fluctuations on a string in a way that effectively suppresses popcorn-type diquark formation.

A final physics modelling option we have implemented is that of Strange Junctions, which allows for the effective string tension near a string junction to be higher than far away from junctions.

A tuning procedure outlined in section 3.2 was carried out, with the strength parameters in each of the closepacking, popcorn destructive interference, and strange-junction models being predominantly tuned to ALICE light hadronic ratios with respect to ALICE charged

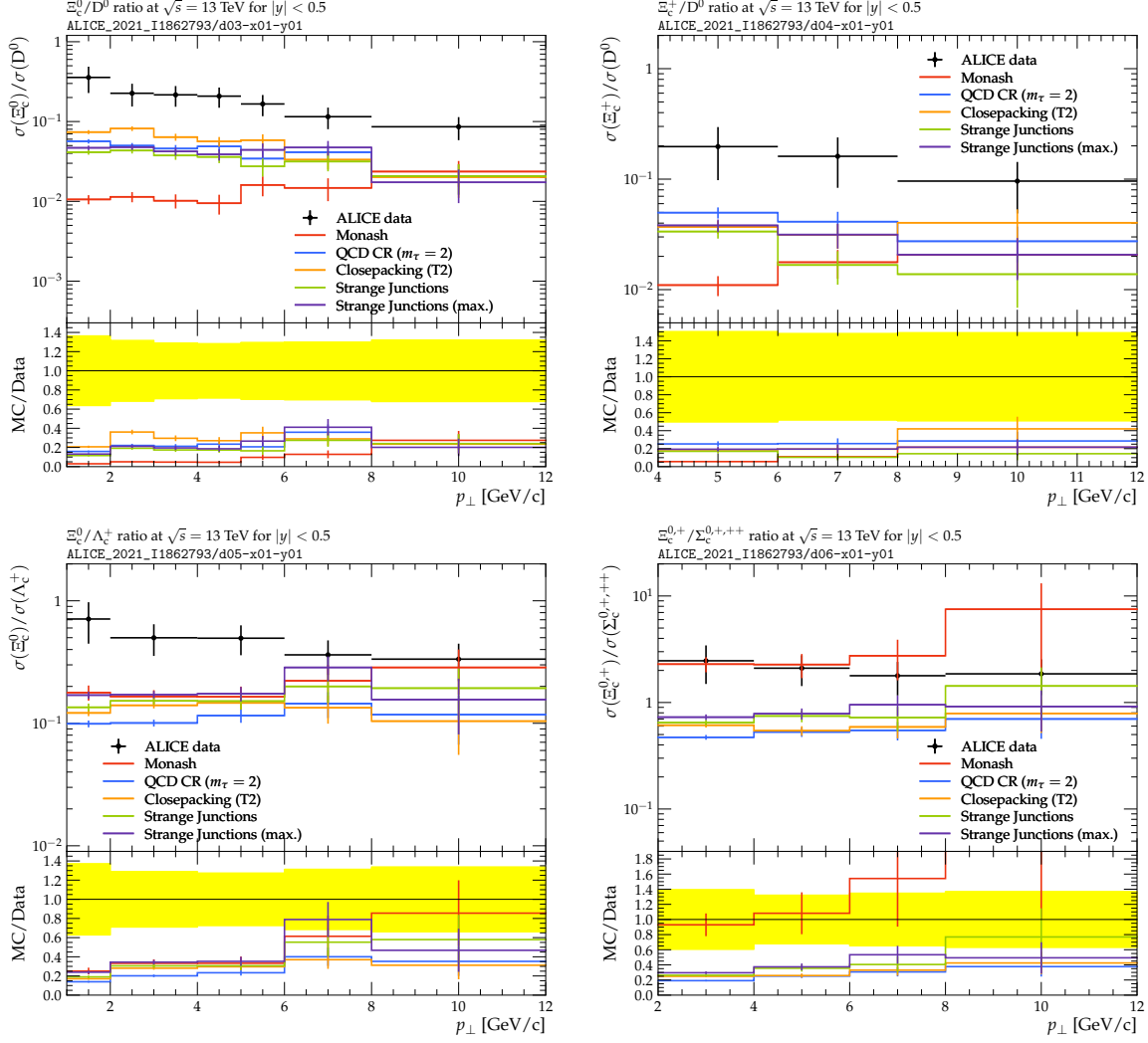


Figure 17. Ratios of Ξ_c to various hadronic species with respect to $p_\perp$, measured by ALICE [70] for $s = 13$ TeV inelastic events. Measurements are conducted at midrapidity, $|y| < 0.5$, for prompt charm hadrons. Top left: Ξ_c^0/D^0 . Top right: Ξ_c^+/D^0 . Bottom left: Ξ_c^0/Λ_c^+ . Bottom right: $\Xi_c^{0,+}/\Sigma_c^{0,+,++}$.

multiplicity classes (see refs. [32, 33]). In particular, three different tuning configurations were considered — Closepacking Trieste 2 (T2) which was a combinatorial tune of all of the three of the above models, Closepacking (simple) which included only closepacking and popcorn destructive interference, and Strange Junctions which exclusively implemented the strange-junction model. These tunes produced varying levels of strangeness-enhancement effects with respect to charged multiplicity, with results summarized in tables 2 and 3. None of the tuning configurations managed to describe all of the observables simultaneously, with particularly interesting disparities occurring for the double-strange ϕ/π ratio and the non-strange baryon-to-meson p/π ratio. Nonetheless, they proved to be competitive with or better than previous modelling attempts, in producing enhanced strange-hadron levels whilst avoiding egregiously overshooting proton production. We attribute this mainly to the

popcorn destructive interference modelling, which — being developed in this work — was not enabled in the existing Monash, QCD CR, and Rope hadronization tunes.

Additional to the light sector, section 3.4 demonstrates how these tunes fare in describing the prompt charm hadronic ratios. Despite their enhanced strangeness production, none of the models we considered are able to reach agreement with the Ξ_c ratios. We find this puzzling, and would recommend further dedicated investigations into Ξ_c and/or other heavy-flavoured strange hadrons.

Another potential future development would be to fully generalize the closepacking model to arbitrary collision processes and ensure applicability to all observables. In the current form, the closepacking and popcorn destructive interference implementations rely on the assumption that the bulk of colour flow is aligned along the z -axis. This beam-axis dependence of course no longer has significance in the case of e^+e^- collision events, nor could it be applied to the description of hadrochemistry in high- $p_\perp$ jets. The study of cleaner string environments such as $e^+e^- \rightarrow WW$ events would provide an extremely useful baseline expectation for the closepacking model. The accurate description of jet hadrochemistry is also vital to pin down, given, e.g., its impact on jet calibrations, cf., e.g., [72].

Acknowledgments

We thank the Rudolf Peierls Centre for Theoretical Physics, University of Oxford, for hospitality and a wonderful working environment during the completion of this work. This work was supported by the Royal Society, grant no. RSWVF-R1-231006, by the Australian Research Council Discovery Project grant no. DP230103014, and by the Monash-Warwick Alliance for Particle Physics. This work was supported also by the Friuli Venezia Giulia region with Microgrant J93C22001380002.

A Diquark scaling for closepacking

The probability `StringFlav:probQQtoQ` scales in a more complicated manner in comparison to the other flavour probabilities given a modified string tension. The probability of diquark compared to quark production from string breaks can be parameterized as,

$$P(qq : q) = \frac{\sum_{qs} \mathcal{P}_{qq_s}}{\sum_q \mathcal{P}_q} = \alpha \frac{\mathcal{P}_{ud0}}{\mathcal{P}_u}, \quad (\text{A.1})$$

with the variable α here being a function of the other flavour probabilities. Given the parameters `StringFlav:probStoUD`, `StringFlav:probSqtoQQ`, `StringFlav:probQQ1toQQ0` being represented by symbols ρ , x , and y respectively, then the factor α can be parameterized as,

$$\alpha = \frac{1 + 2x\rho + 9y + 6x\rho y + 3yx^2\rho^2}{2 + \rho}. \quad (\text{A.2})$$

In the parameterization of $P'(qq : q)$ given in eq. (2.12), α and α' take the same form as eq. (A.2), simply with modified flavour probabilities defined by κ_0 and κ_{eff} respectively. Important to note is that this scaling of the diquark probability works under the assumption that an effective string tension does not effect the probability $\mathcal{P}_{ud0}/\mathcal{P}_u$, and assumes Schwinger-type diquark-antidiquark pair production.

# of overlapping strings $C_1 = p + q$	# parallel p	# anti- parallel q	$\frac{\kappa_p}{\kappa_0}$ $\frac{6+2p+q}{8}$	$\frac{\kappa_q}{\kappa_0}$ $\frac{6+2q+p}{8}$	Total $\frac{\kappa}{\kappa_0}$ $\frac{p\kappa_p + q\kappa_q}{\kappa_0}$	Casimir ratio $\frac{C_2(p,q)}{C_2(1,0)}$	SU(3) Rep $\frac{(p+1)(q+1)(p+q+2)}{2}$
1	1	0	1	—	1	1	3
2	2	0	1.25	—	2.5	2.5	6
2	1	1	1.125	1.125	2.25	2.25	8
3	3	0	1.5	—	4.5	4.5	10
3	2	1	1.375	1.25	4	4	15
4	4	0	1.75	—	7	7	15'
4	3	1	1.625	1.375	6.25	6.25	24
4	2	2	1.5	1.5	6	6	27
5	5	0	2	—	10	10	21
5	4	1	1.875	1.5	9	9	35
5	3	2	1.75	1.625	8.5	8.5	42
6	6	0	2.25	—	13.5	13.5	28
6	5	1	2.125	1.625	12.25	12.25	48
6	4	2	2	1.75	11.5	11.5	60
6	3	3	1.875	1.875	11.25	11.25	64
7	7	0	2.5	—	17.5	17.5	36
7	6	1	2.375	1.75	16	16	63
7	5	2	2.25	1.875	15	15	81
7	4	3	2.125	2	14.5	14.5	90
8	8	0	2.75	—	22	22	45
8	7	1	2.625	1.875	20.25	20.25	80
8	6	2	2.5	2	19	19	105
8	5	3	2.375	2.125	18.25	18.25	120
8	4	4	2.25	2.25	18	18	125

Table 5. Numerical values of the string-tension enhancement factors κ_p/κ_0 and κ_q/κ_0 , for closepacking with Casimir scaling, for up to 8 total strings, with p parallel flux directions and q antiparallel ones, in the absence of suppression factors. The summed string tensions, κ/κ_0 , agree with the exact ratios of SU(3) quadratic Casimirs $C_2(p,q)/C_2(1,0)$. The corresponding SU(3) multiplets are also given. Note that there are two different 15-plets, labelled **15** and **15'** respectively.

B Table of Casimir-scaling coefficients

Table 5 shows the baseline string-tension enhancement factors κ_p/κ_0 for each string in an environment with p total strings with parallel flux orientations (including the string itself) and q antiparallel ones, in the absence of suppression factors. (The numbers for $q > p$ are identical, with $p \leftrightarrow q$, and hence are not shown.) Also shown is the total (summed) string tension normalized to the tension for a single string, which agrees with the ratio of the quadratic Casimir in SU(3), $C_2(p,q)$, to the fundamental one, $C_2(1,0) = C_F = 4/3$. The last column gives the size of the corresponding multiplets in SU(3).

	Monash	QCD CR ($m_\tau = 2$)	Rope Hadroni- zation	Close- packing (T1)
StringZ:useOldAExtra	= off	= off	= off	= off
StringPT:sigma	= 0.335	= 0.335	= 0.335	= 0.335
StringZ:aLund	= 0.68	= 0.36	= 0.68	= 0.68
StringZ:bLund	= 0.98	= 0.56	= 0.98	= 0.98
StringFlav:probQQtoQ	= 0.081	= 0.078	= 0.081	= 0.081
StringFlav:probStoUD	= 0.217	= 0.2	= 0.217	= 0.217
StringFlav:probQQ1toQQ0join		= {0.5, 0.7, 0.9, 1.0}		
MultiPartonInteractions:pT0Ref	= 2.28	= 2.15	= 2.15	= 2.194
BeamRemnants:remnantMode	= 0	= 1	= 1	= 1
BeamRemnants:saturation	—	= 5	= 5	= 5
ColourReconnection:mode	= 0	= 1	= 1	= 1
ColourReconnection:timeDilationMode	—	= 2	= 2	= 2
ColourReconnection:allowDoubleJunRem	—	= off	= off	= off
ColourReconnection:allowJunctions	—	= on	= on	= on
ColourReconnection:mPseudo	—	= 0.3	= 0.3	= 0.403
ColourReconnection:m0	—	= 0.3	= 0.3	= 0.618
ColourReconnection:junctionCorrection	—	= 1.20	= 1.20	= 1.349
ColourReconnection:timeDilationPar	—	= 0.18	= 0.18	= 0.18
ClosePacking:doClosePacking	= off	= off	= off	= on
ClosePacking:enhanceStrange	—	—	—	= 0.014
ClosePacking:enhancePT	—	—	—	= 0.014
ClosePacking:PT0	—	—	—	= 2.0
ClosePacking:doEnhanceDiquark	= off	= off	= off	= off
StringFragmentation:doStrangeJunctions	= off	= off	= off	= on
StringFragmentation:enhanceStrangeJunction	—	—	—	= 0.540
ClosePacking:baryonSup	—	—	—	= 0.928
ClosePacking:parallelBaryonSup	—	—	—	= 0
Ropewalk:RopeHadronization	= off	= off	= on	= off
Ropewalk:doShoving	—	—	= on	—
Ropewalk:tInit	—	—	= 1.5	—
Ropewalk:deltat	—	—	= 0.05	—
Ropewalk:tShove	—	—	= 0.1	—
Ropewalk:gAmplitude	—	—	= 0.	—
Ropewalk:doFlavour	—	—	= on	—
Ropewalk:r0	—	—	= 0.5	—
Ropewalk:m0	—	—	= 0.2	—
Ropewalk:beta	—	—	= 0.1	—
PartonVertex:setVertex	—	—	= on	—
PartonVertex:protonRadius	—	—	= 0.7	—
PartonVertex:emissionWidth	—	—	= 0.1	—
ParticleDecays:limitTau0	= on	= on	= on	= on
ParticleDecays:tau0Max	= 10	= 10	= 10	= 10

Table 6. Parameter values for the following tunes: Monash, QCD CR ($m_\tau = 2$), Rope Hadronization and Closepacking (T1). Parameters for the new tunes developed in this work are listed in table 1.

C Reference tune parameters

Here, we list the parameter sets of the existing models and tunes that we have compared to in this work, including the Monash tune [11], QCD CR ($m_\tau = 2$) [45], Rope Hadronization, and Closepacking (T1) [60]. Note that older tunes normally use `StringZ:useOldAExtra = on`, to reproduce a bug in the versions they were tuned with, while the newer `off` setting was used for this work (without retuning the fragmentation parameters).

D Further plots

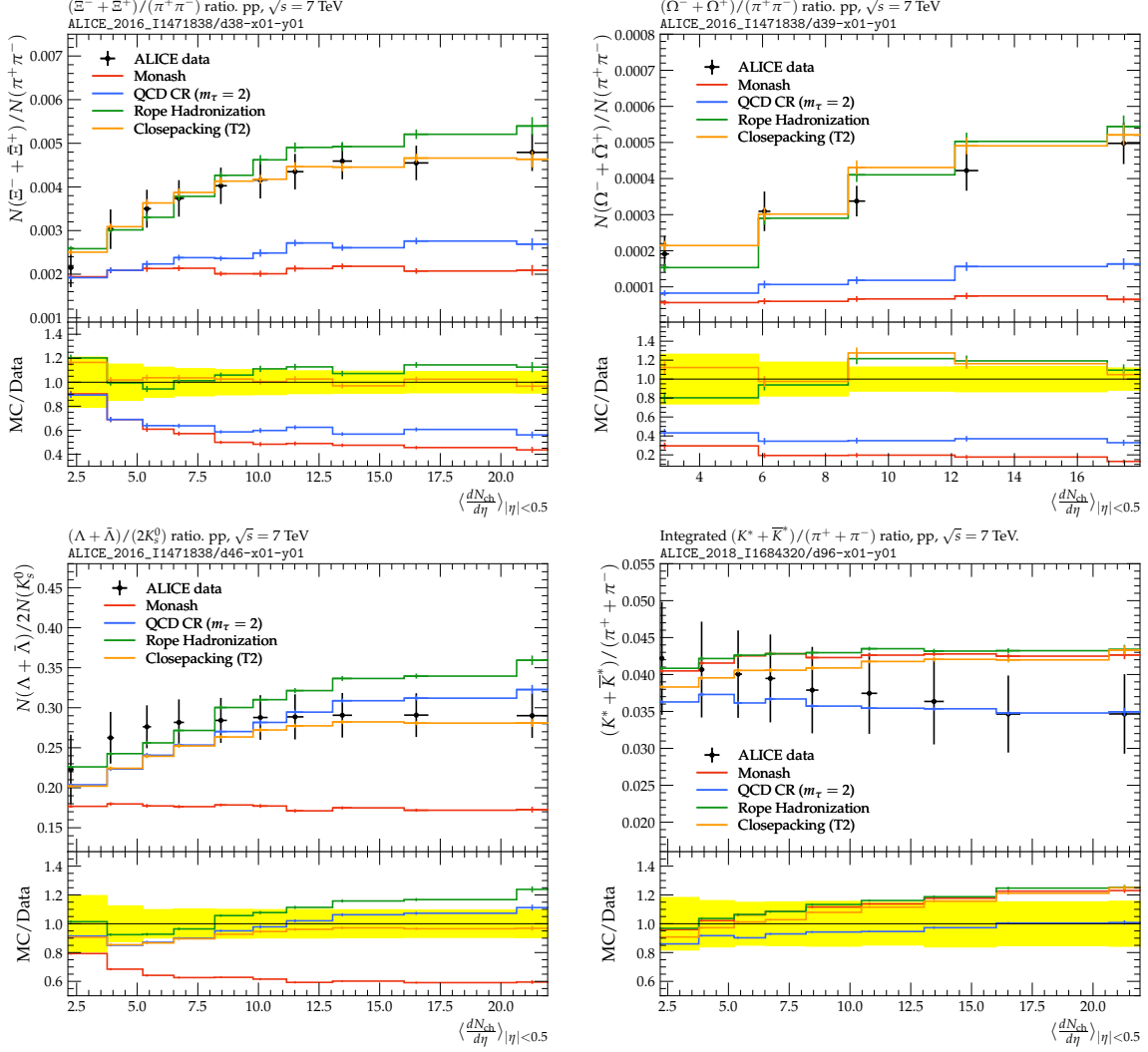


Figure 18. Hadron-to-pion ratios with respect to ALICE midrapidity charged multiplicity classes. PYTHIA simulations for parameters in table 6 of 7 TeV INEL>0 events are run with a lifetime cut of $\tau_{\text{max}} = 10\text{mm}/c$, and no $p_{\perp}$ cuts on final-state particles. Top left: Ξ/π [32]. Top right: Ω/π [32]. Bottom left: $\Lambda/2K_S^0$ [32]. Bottom right: K^*/π [33].

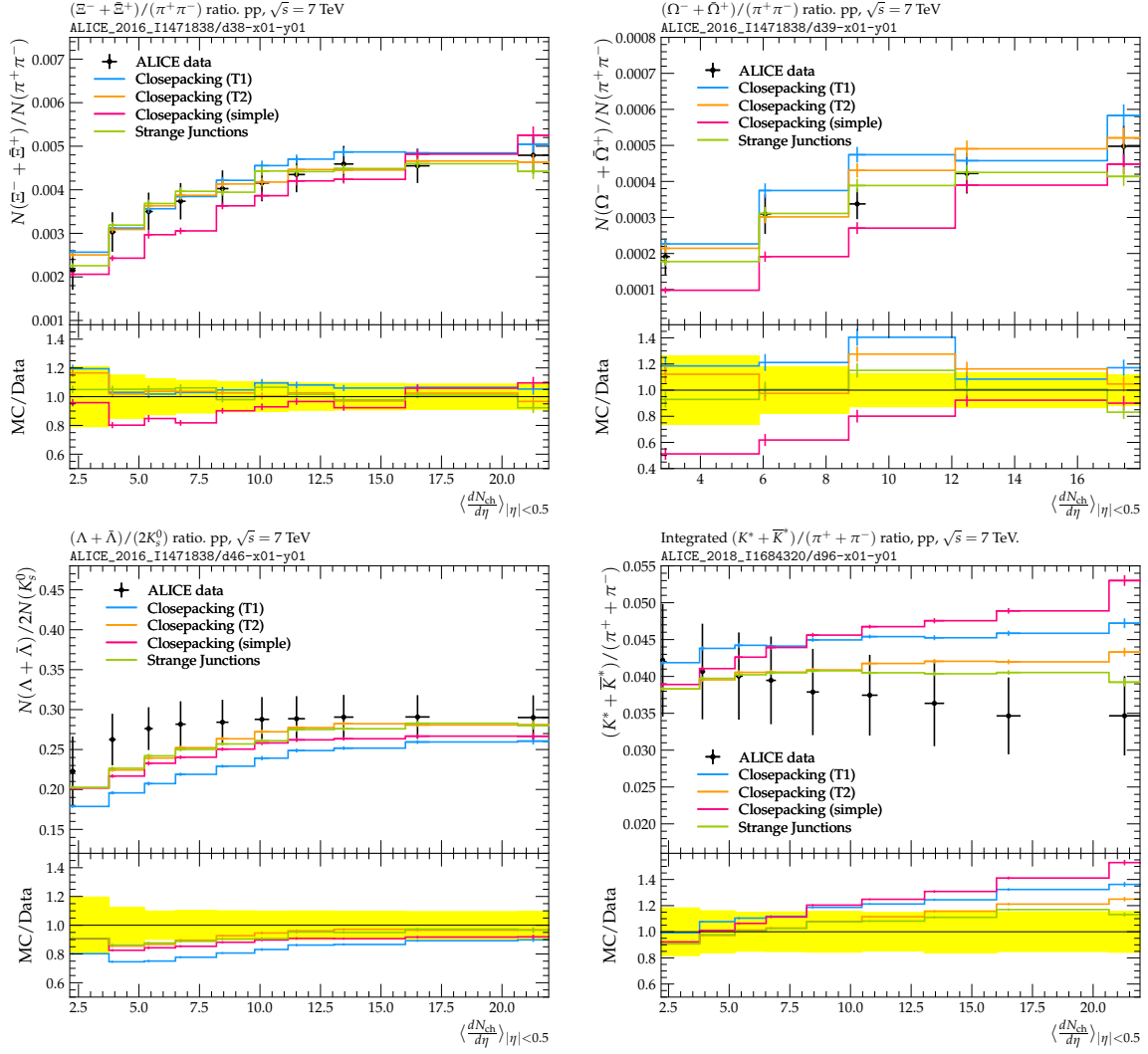


Figure 19. Hadron-to-pion ratios with respect to ALICE midrapidity charged multiplicity classes. PYTHIA simulations for parameters in table 1 of 7 TeV INEL>0 events are run with a lifetime cut of $\tau_{\max} = 10\text{mm}/c$, and no $p_{\perp}$ cuts on final-state particles. Top left: Ξ/π [32]. Top right: Ω/π [32]. Bottom left: $\Lambda/2K_S^0$ [32]. Bottom right: K^*/π [33].

E Colour map for goodness-of-fit assessment

Table 7 shows the colour scheme, based on χ_{GoF}^2 values, adopted in this work.

Colour	Hex Code	σ_{GoF}	χ_{GoF}^2	p_{GoF}
	#40B748	0.00 — 0.33	0.00 — 0.11	0.74 — 1.00
	#56BB46	0.33 — 0.67	0.11 — 0.44	0.50 — 0.74
	#6CBD45	0.67 — 1.00	0.44 — 1.00	0.32 — 0.50
	#81C342	1.00 — 1.33	1.00 — 1.78	0.18 — 0.32
	#BDCD30	1.33 — 1.67	1.78 — 2.78	0.096 — 0.18
	#FDCD06	1.67 — 2.00	2.78 — 4.00	0.046 — 0.096
	#FDB314	2.00 — 2.33	4.00 — 5.44	0.020 — 0.046
	#F79717	2.33 — 2.67	5.44 — 7.11	0.0077 — 0.020
	#F78220	2.67 — 3.00	7.11 — 9.00	0.0027 — 0.0077
	#F16820	3.00 — 3.33	9.00 — 11.11	$[0.86 — 2.7] \times 10^{-3}$
	#EF4A24	3.33 — 3.67	11.11 — 13.44	$[0.25 — 0.86] \times 10^{-3}$
	#FF0000	3.67 — 4.00	13.44 — 16.00	$[0.63 — 2.5] \times 10^{-4}$
	#E63B7A	4.00 — 4.33	16.00 — 18.78	$[1.5 — 6.3] \times 10^{-5}$
	#B92D5D	4.33 — 4.67	18.78 — 21.78	$[0.3 — 1.5] \times 10^{-5}$
	#9A244F	4.67 — 5.00	21.78 — 25.00	$[1 — 3] \times 10^{-6}$
	#000000	≥ 5.00	≥ 25.00	$\leq 10^{-6}$

Table 7. Reduced χ_{GoF}^2 colour map for Goodness of Fit (GoF) assessment, with equivalent standard deviations $\sigma \equiv \sqrt{\chi_{\text{GoF}}^2}$, and p values (for a normal distribution) indicating the probability that a purely stochastic fluctuation could yield a higher χ_{GoF}^2 .

F Tuning observables

The observables used for tuning, along with their assigned weights, order of polynomial, and versions of PYTHIA used are listed in this section, for each tuning step. The Rivet analyses utilized for the tuning include:⁶

- ALICE_2016_I1471838 [32]
- ALICE_2018_I1684320⁷ [33]
- ATLAS_2010_S8918562 [73]
- CMS_2010_S8656010 [74]
- CMS_2011_S8884919 [65]
- CMS_2011_S8978280 [49]
- LHCb_2011_I917009 [75]

⁶Rivet 3 naming conventions.

⁷Note that this Rivet analysis is not yet published and publicly available.

Rivet path	Description	Axis	Weight
/ATLAS_2010_S8918562/d03-x01-y01	Charged particle η at 7 TeV, track $p_{\perp} > 500$ MeV, for $N_{\text{ch}} \geq 1$	y-axis: $1/N_{\text{ev}} dN_{\text{ch}}/d\eta$ x-axis: η	1.0
/ATLAS_2010_S8918562/d05-x01-y01	Charged particle η at 7 TeV, track $p_{\perp} > 100$ MeV, for $N_{\text{ch}} \geq 2$	y-axis: $1/N_{\text{ev}} dN_{\text{ch}}/d\eta$ x-axis: η	1.0
/ATLAS_2010_S8918562/d07-x01-y01	Charged particle η at 7 TeV, track $p_{\perp} > 500$ MeV, for $N_{\text{ch}} \geq 6$	y-axis: $1/N_{\text{ev}} dN_{\text{ch}}/d\eta$ x-axis: η	1.0
/ATLAS_2010_S8918562/d23-x01-y01	Charged $\langle p_{\perp} \rangle$ vs. N_{ch} at 7 TeV, track $p_{\perp} > 500$ MeV, for $N_{\text{ch}} \geq 1$	y-axis: $\langle p_{\perp} \rangle$ [GeV] x-axis: N_{ch}	1.0
/ATLAS_2010_S8918562/d25-x01-y01	Charged $\langle p_{\perp} \rangle$ vs. N_{ch} at 7 TeV, track $p_{\perp} > 100$ MeV, for $N_{\text{ch}} \geq 2$	y-axis: $\langle p_{\perp} \rangle$ [GeV] x-axis: N_{ch}	1.0
/ATLAS_2010_S8918562/d27-x01-y01	Charged particle η at 7 TeV, track $p_{\perp} > 100$ MeV, for $N_{\text{ch}} \geq 20$	y-axis: $1/N_{\text{ev}} dN_{\text{ch}}/d\eta$ x-axis: η	1.0

Table 8. Observables and corresponding weights used in the first tuning step.

F.1 Tuning step 1

Weights were set to 1 for all observables used in the first tuning step, as reported in table 8. Simulations were run using PYTHIA 8.315 with an unpublished closepacking code update, and 4th order polynomials were used for interpolation.

F.2 Tuning step 2

Weights were set to 1 for all observables used in the second tuning step, as reported in table 9. Other configurations were also tested. For instance, in one test, a weight of 2 was assigned only to the ϕ/π ratio. In another configuration, the same weight was applied to ϕ/π , K^{*0}/π and p/π . However, the configuration with all weights set to one ultimately performed best in terms of resulting χ^2_{GoF} values. Simulations were run using PYTHIA 8.316, using 5th order polynomials for interpolation.

Rivet path	Description	Axis	Weight
/ALICE_2016_I1471838/d36-x01-y01	$2K_s^0/(\pi^+\pi^-)$ ratio	y-axis: $2N(K_s^0)/N(\pi^+\pi^-)$ x-axis: $\langle \frac{dN_{\text{ch}}}{d\eta} \rangle_{ \eta <0.5}$	1.0
/ALICE_2016_I1471838/d37-x01-y01	$(\Lambda + \bar{\Lambda})/(\pi^+\pi^-)$ ratio	y-axis: $N(\Lambda + \bar{\Lambda})/N(\pi^+\pi^-)$ x-axis: $\langle \frac{dN_{\text{ch}}}{d\eta} \rangle_{ \eta <0.5}$	1.0
/ALICE_2016_I1471838/d38-x01-y01	$(\Xi^- + \bar{\Xi}^+)/(\pi^+\pi^-)$ ratio	y-axis: $N(\Xi^- + \bar{\Xi}^+)/N(\pi^+\pi^-)$ x-axis: $\langle \frac{dN_{\text{ch}}}{d\eta} \rangle_{ \eta <0.5}$	1.0
/ALICE_2016_I1471838/d39-x01-y01	$(\Omega^- + \bar{\Omega}^+)/(\pi^+\pi^-)$ ratio	y-axis: $N(\Omega^- + \bar{\Omega}^+)/N(\pi^+\pi^-)$ x-axis: $\langle \frac{dN_{\text{ch}}}{d\eta} \rangle_{ \eta <0.5}$	1.0
/ALICE_2016_I1471838/d46-x01-y01	$(\Lambda + \bar{\Lambda})/(2K_s^0)$ ratio	y-axis: $N(\Lambda + \bar{\Lambda})/2N(K_s^0)$ x-axis: $\langle \frac{dN_{\text{ch}}}{d\eta} \rangle_{ \eta <0.5}$	1.0
/ALICE_2016_I1471838/d47-x01-y01	$(p + \bar{p})/(\pi^+ + \pi^-)$ ratio	y-axis: $N(p + \bar{p})/N(\pi^+ + \pi^-)$ x-axis: $\langle \frac{dN_{\text{ch}}}{d\eta} \rangle_{ \eta <0.5}$	1.0
/ALICE_2018_I1684320/d96-x01-y01	Integrated $(K^* + \bar{K}^*)/(\pi^+ + \pi^-)$ ratio, pp, $\sqrt{s} = 7$ TeV.	y-axis: $(K^* + \bar{K}^*)/(\pi^+ + \pi^-)$ x-axis: $\langle \frac{dN_{\text{ch}}}{d\eta} \rangle_{ \eta <0.5}$	1.0

Continued on next page

Rivet path	Description	Axis	Weight
/ALICE_2018_I1684320/d97-x01-y01	Integrated $\phi/(\pi^+ + \pi^-)$ ratio, pp, $\sqrt{s} = 7$ TeV.	y-axis: $\phi/(\pi^+ + \pi^-)$ x-axis: $\langle \frac{dN_{ch}}{d\eta} \rangle_{ \eta < 0.5}$	1.0
/CMS_2011_S8978280/d05-x01-y02	Ξ^- rapidity distribution at $\sqrt{s} = 7$ TeV	y-axis: $(1/N_{NSD}) dN/dy$ x-axis: $\Xi^- y $	1.0
/CMS_2011_S8978280/d09-x01-y02	Λ/K_S^0 versus rapidity at $\sqrt{s} = 7$ TeV	y-axis: $N(\Lambda) / N(K_S^0)$ x-axis: $ y $	1.0
/CMS_2011_S8978280/d10-x01-y02	Ξ^-/Λ versus rapidity at $\sqrt{s} = 7$ TeV	y-axis: $N(\Xi^-) / N(\Lambda)$ x-axis: $ y $	1.0
/LHCB_2011_I917009/d12-x01-y01	$\bar{\Lambda}/K_s^0$ ratio at $\sqrt{s} = 7$ TeV ($0.15 < p_\perp < 2.5$ GeV/c)	y-axis: $\bar{\Lambda}/K_s^0$ x-axis: Rapidity (y)	1.0
/LHCB_2011_I917009/d16-x01-y01	$\bar{\Lambda}/K_s^0$ ratio at $\sqrt{s} = 7$ TeV ($0.15 < p_\perp < 2.5$ GeV/c)	y-axis: $\bar{\Lambda}/K_s^0$ x-axis: Rapidity loss (Δy)	1.0

End of table.

Table 9. Observables and corresponding weights used in the second tuning step.

F.3 Tuning step 3

In this third step, weights differing from unity are used to account for the different number of observables in each category. In particular, weight 6 is assigned to pseudorapidity distributions, and weight 8 is assigned to the $\langle p_\perp \rangle$ vs n_{ch} distributions, as reported in table 10. Simulations were run using PYTHIA 8.316, using 5th order polynomials for interpolation.

Rivet path	Description	Axis	Weight
/ALICE_2016_I1471838/d01-x01-y01	$K_s^0, p_\perp$ spectrum V0M Class I	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d02-x01-y01	$K_s^0, p_\perp$ spectrum V0M Class II	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d03-x01-y01	$K_s^0, p_\perp$ spectrum V0M Class III	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d04-x01-y01	$K_s^0, p_\perp$ spectrum V0M Class IV	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d05-x01-y01	$K_s^0, p_\perp$ spectrum V0M Class V	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d06-x01-y01	$K_s^0, p_\perp$ spectrum V0M Class VI	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d07-x01-y01	$K_s^0, p_\perp$ spectrum V0M Class VII	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d08-x01-y01	$K_s^0, p_\perp$ spectrum V0M Class VIII	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d09-x01-y01	$K_s^0, p_\perp$ spectrum V0M Class IX	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0

Continued on next page

Rivet path	Description	Axis	Weight
/ALICE_2016_I1471838/d10-x01-y01	$K_s^0, p_\perp$ spectrum V0M Class X	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d11-x01-y01	$\Lambda + \bar{\Lambda}, p_\perp$ spectrum V0M Class I	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d12-x01-y01	$\Lambda + \bar{\Lambda}, p_\perp$ spectrum V0M Class II	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d13-x01-y01	$\Lambda + \bar{\Lambda}, p_\perp$ spectrum V0M Class III	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d14-x01-y01	$\Lambda + \bar{\Lambda}, p_\perp$ spectrum V0M Class IV	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d15-x01-y01	$\Lambda + \bar{\Lambda}, p_\perp$ spectrum V0M Class V	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d16-x01-y01	$\Lambda + \bar{\Lambda}, p_\perp$ spectrum V0M Class VI	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d17-x01-y01	$\Lambda + \bar{\Lambda}, p_\perp$ spectrum V0M Class VII	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d18-x01-y01	$\Lambda + \bar{\Lambda}, p_\perp$ spectrum V0M Class VIII	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d19-x01-y01	$\Lambda + \bar{\Lambda}, p_\perp$ spectrum V0M Class IX	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d20-x01-y01	$\Lambda + \bar{\Lambda}, p_\perp$ spectrum V0M Class X	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d21-x01-y01	$\Xi^- + \bar{\Xi}^+, p_\perp$ spectrum V0M Class I	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d22-x01-y01	$\Xi^- + \bar{\Xi}^+, p_\perp$ spectrum V0M Class II	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d23-x01-y01	$\Xi^- + \bar{\Xi}^+, p_\perp$ spectrum V0M Class III	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d24-x01-y01	$\Xi^- + \bar{\Xi}^+, p_\perp$ spectrum V0M Class IV	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d25-x01-y01	$\Xi^- + \bar{\Xi}^+, p_\perp$ spectrum V0M Class V	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d26-x01-y01	$\Xi^- + \bar{\Xi}^+, p_\perp$ spectrum V0M Class VI	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d27-x01-y01	$\Xi^- + \bar{\Xi}^+, p_\perp$ spectrum V0M Class VII	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d28-x01-y01	$\Xi^- + \bar{\Xi}^+, p_\perp$ spectrum V0M Class VIII	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d29-x01-y01	$\Xi^- + \bar{\Xi}^+, p_\perp$ spectrum V0M Class IX	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0

Continued on next page

Rivet path	Description	Axis	Weight
/ALICE_2016_I1471838/d30-x01-y01	$\Xi^- + \bar{\Xi}^+, p_\perp$ spectrum V0M Class X	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d31-x01-y01	$\Omega^- + \bar{\Omega}^+, p_\perp$ spectrum V0M Class I+II	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d32-x01-y01	$\Omega^- + \bar{\Omega}^+, p_\perp$ spectrum V0M Class III+IV	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d33-x01-y01	$\Omega^- + \bar{\Omega}^+, p_\perp$ spectrum V0M Class V+VI	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d34-x01-y01	$\Omega^- + \bar{\Omega}^+, p_\perp$ spectrum V0M Class VI+VIII	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2016_I1471838/d35-x01-y01	$\Omega^- + \bar{\Omega}^+, p_\perp$ spectrum V0M Class IX+X	y-axis: $1/N \frac{d^2 N}{dy dp_\perp}$ x-axis: $p_\perp$ [GeV]	1.0
/ALICE_2018_I1684320/d01-x01-y01	Charged particles, $p_\perp$ spectra, $ \eta < 0.5$, V0M Class I, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{d\eta dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d02-x01-y01	Charged particles, $p_\perp$ spectra, $ \eta < 0.5$, V0M Class II, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{d\eta dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d03-x01-y01	Charged particles, $p_\perp$ spectra, $ \eta < 0.5$, V0M Class III, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{d\eta dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d04-x01-y01	Charged particles, $p_\perp$ spectra, $ \eta < 0.5$, V0M Class IV, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{d\eta dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d05-x01-y01	Charged particles, $p_\perp$ spectra, $ \eta < 0.5$, V0M Class V, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{d\eta dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d06-x01-y01	Charged particles, $p_\perp$ spectra, $ \eta < 0.5$, V0M Class VI, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{d\eta dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d07-x01-y01	Charged particles, $p_\perp$ spectra, $ \eta < 0.5$, V0M Class VII, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{d\eta dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d08-x01-y01	Charged particles, $p_\perp$ spectra, $ \eta < 0.5$, V0M Class VIII, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{d\eta dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d09-x01-y01	Charged particles, $p_\perp$ spectra, $ \eta < 0.5$, V0M Class IX, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{d\eta dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d10-x01-y01	Charged particles, $p_\perp$ spectra, $ \eta < 0.5$, V0M Class X, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{d\eta dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0

Continued on next page

Rivet path	Description	Axis	Weight
/ALICE_2018_I1684320/d11-x01-y01	$\pi^+ + \pi^-$, $p_\perp$ spectra, V0M Class I, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d12-x01-y01	$\pi^+ + \pi^-$, $p_\perp$ spectra, V0M Class II, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d13-x01-y01	$\pi^+ + \pi^-$, $p_\perp$ spectra, V0M Class III, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d14-x01-y01	$\pi^+ + \pi^-$, $p_\perp$ spectra, V0M Class IV, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d15-x01-y01	$\pi^+ + \pi^-$, $p_\perp$ spectra, V0M Class V, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d16-x01-y01	$\pi^+ + \pi^-$, $p_\perp$ spectra, V0M Class VI, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d17-x01-y01	$\pi^+ + \pi^-$, $p_\perp$ spectra, V0M Class VII, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d18-x01-y01	$\pi^+ + \pi^-$, $p_\perp$ spectra, V0M Class VIII, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d19-x01-y01	$\pi^+ + \pi^-$, $p_\perp$ spectra, V0M Class IX, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d20-x01-y01	$\pi^+ + \pi^-$, $p_\perp$ spectra, V0M Class X, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d21-x01-y01	$K^+ + K^-$, $p_\perp$ spectra, V0M Class I, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d22-x01-y01	$K^+ + K^-$, $p_\perp$ spectra, V0M Class II, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d23-x01-y01	$K^+ + K^-$, $p_\perp$ spectra, V0M Class III, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d24-x01-y01	$K^+ + K^-$, $p_\perp$ spectra, V0M Class IV, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d25-x01-y01	$K^+ + K^-$, $p_\perp$ spectra, V0M Class V, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) $^{-1}$] x-axis: $p_\perp$ [GeV/c]	1.0

Continued on next page

Rivet path	Description	Axis	Weight
/ALICE_2018_I1684320/d26-x01-y01	$K^+ + K^-$, $p_\perp$ spectra, V0M Class VI, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d27-x01-y01	$K^+ + K^-$, $p_\perp$ spectra, V0M Class VII, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d28-x01-y01	$K^+ + K^-$, $p_\perp$ spectra, V0M Class VIII, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d29-x01-y01	$K^+ + K^-$, $p_\perp$ spectra, V0M Class IX, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d30-x01-y01	$K^+ + K^-$, $p_\perp$ spectra, V0M Class X, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d31-x01-y01	$p + \bar{p}$, $p_\perp$ spectra, V0M Class I, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d32-x01-y01	$p + \bar{p}$, $p_\perp$ spectra, V0M Class II, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d33-x01-y01	$p + \bar{p}$, $p_\perp$ spectra, V0M Class III, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d34-x01-y01	$p + \bar{p}$, $p_\perp$ spectra, V0M Class VI, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d35-x01-y01	$p + \bar{p}$, $p_\perp$ spectra, V0M Class V, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d36-x01-y01	$p + \bar{p}$, $p_\perp$ spectra, V0M Class VI, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d37-x01-y01	$p + \bar{p}$, $p_\perp$ spectra, V0M Class VII, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d38-x01-y01	$p + \bar{p}$, $p_\perp$ spectra, V0M Class VIII, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d39-x01-y01	$p + \bar{p}$, $p_\perp$ spectra, V0M Class IX, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d40-x01-y01	$p + \bar{p}$, $p_\perp$ spectra, V0M Class X, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0

Continued on next page

Rivet path	Description	Axis	Weight
/ALICE_2018_I1684320/d41-x01-y01	$K^* + \bar{K}^*$, $p_\perp$ spectra, V0M Class I, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d42-x01-y01	$K^* + \bar{K}^*$, $p_\perp$ spectra, V0M Class II, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d43-x01-y01	$K^* + \bar{K}^*$, $p_\perp$ spectra, V0M Class II, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d44-x01-y01	$K^* + \bar{K}^*$, $p_\perp$ spectra, V0M Class IV+V, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d45-x01-y01	$K^* + \bar{K}^*$, $p_\perp$ spectra, V0M Class VI, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d46-x01-y01	$K^* + \bar{K}^*$, $p_\perp$ spectra, V0M Class VII, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d47-x01-y01	$K^* + \bar{K}^*$, $p_\perp$ spectra, V0M Class VIII, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d48-x01-y01	$K^* + \bar{K}^*$, $p_\perp$ spectra, V0M Class IX, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d49-x01-y01	$K^* + \bar{K}^*$, $p_\perp$ spectra, V0M Class X, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d50-x01-y01	ϕ , $p_\perp$ spectra, V0M Class I, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d51-x01-y01	ϕ , $p_\perp$ spectra, V0M Class II, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d52-x01-y01	ϕ , $p_\perp$ spectra, V0M Class III, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d53-x01-y01	ϕ , $p_\perp$ spectra, V0M Class IV+V, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d54-x01-y01	ϕ , $p_\perp$ spectra, V0M Class VI, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0
/ALICE_2018_I1684320/d55-x01-y01	ϕ , $p_\perp$ spectra, V0M Class VII, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_\perp}$ [(GeV/c) ⁻¹] x-axis: $p_\perp$ [GeV/c]	1.0

Continued on next page

Rivet path	Description	Axis	Weight
/ALICE_2018_I1684320/d56-x01-y01	ϕ , $p_{\perp}$ spectra, V0M Class VIII, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_{\perp}}$ [(GeV/c) ⁻¹] x-axis: $p_{\perp}$ [GeV/c]	1.0
/ALICE_2018_I1684320/d57-x01-y01	ϕ , $p_{\perp}$ spectra, V0M Class IX, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_{\perp}}$ [(GeV/c) ⁻¹] x-axis: $p_{\perp}$ [GeV/c]	1.0
/ALICE_2018_I1684320/d58-x01-y01	ϕ , $p_{\perp}$ spectra, V0M Class X, pp, $\sqrt{s} = 7$ TeV	y-axis: $1/N_{\text{evt}} \frac{d^2 N}{dy dp_{\perp}}$ [(GeV/c) ⁻¹] x-axis: $p_{\perp}$ [GeV/c]	1.0
/ATLAS_2010_S8918562/d03-x01-y01	Charged particle η at 7 TeV, track $p_{\perp} > 500$ MeV, for $N_{\text{ch}} \geq 1$	y-axis: $1/N_{\text{ev}} dN_{\text{ch}}/d\eta$ x-axis: η	6.0
/ATLAS_2010_S8918562/d05-x01-y01	Charged particle η at 7 TeV, track $p_{\perp} > 100$ MeV, for $N_{\text{ch}} \geq 2$	y-axis: $1/N_{\text{ev}} dN_{\text{ch}}/d\eta$ x-axis: η	6.0
/ATLAS_2010_S8918562/d07-x01-y01	Charged particle η at 7 TeV, track $p_{\perp} > 500$ MeV, for $N_{\text{ch}} \geq 6$	y-axis: $1/N_{\text{ev}} dN_{\text{ch}}/d\eta$ x-axis: η	6.0
/ATLAS_2010_S8918562/d10-x01-y01	Charged particle $p_{\perp}$ at 7 TeV, track $p_{\perp} > 500$ MeV, for $N_{\text{ch}} \geq 1$	y-axis: $1/N_{\text{ev}} 1/2\pi p_{\perp} d\sigma/d\eta dp_{\perp}$ x-axis: $p_{\perp}$ [GeV]	1.0
/ATLAS_2010_S8918562/d12-x01-y01	Charged particle $p_{\perp}$ at 7 TeV, track $p_{\perp} > 100$ MeV, for $N_{\text{ch}} \geq 2$	y-axis: $1/N_{\text{ev}} 1/2\pi p_{\perp} d\sigma/d\eta dp_{\perp}$ x-axis: $p_{\perp}$ [GeV]	1.0
/ATLAS_2010_S8918562/d14-x01-y01	Charged particle $p_{\perp}$ at 7 TeV, track $p_{\perp} > 500$ MeV, for $N_{\text{ch}} \geq 6$	y-axis: $1/N_{\text{ev}} 1/2\pi p_{\perp} d\sigma/d\eta dp_{\perp}$ x-axis: $p_{\perp}$ [GeV]	1.0
/ATLAS_2010_S8918562/d23-x01-y01	Charged $\langle p_{\perp} \rangle$ vs. N_{ch} at 7 TeV, track $p_{\perp} > 500$ MeV, for $N_{\text{ch}} \geq 1$	y-axis: $\langle p_{\perp} \rangle$ [GeV] x-axis: N_{ch}	8.0
/ATLAS_2010_S8918562/d25-x01-y01	Charged $\langle p_{\perp} \rangle$ vs. N_{ch} at 7 TeV, track $p_{\perp} > 100$ MeV, for $N_{\text{ch}} \geq 2$	y-axis: $\langle p_{\perp} \rangle$ [GeV] x-axis: N_{ch}	8.0
/CMS_2010_S8656010/d01-x01-y03	Charged hadron $p_{\perp}$ for $ \eta = 0.5$ at $\sqrt{s} = 7$ TeV	y-axis: $d^2 N_{\text{ch}}/d\eta dp_{\perp}$ [(GeV/c) ⁻¹] x-axis: $p_{\perp}$ [GeV/c]	1.0
/CMS_2010_S8656010/d01-x01-y04	Charged hadron $p_{\perp}$ for $ \eta = 0.7$ at $\sqrt{s} = 7$ TeV	y-axis: $d^2 N_{\text{ch}}/d\eta dp_{\perp}$ [(GeV/c) ⁻¹] x-axis: $p_{\perp}$ [GeV/c]	1.0
/CMS_2010_S8656010/d02-x01-y01	Charged hadron $p_{\perp}$ for $ \eta = 0.9$ at $\sqrt{s} = 7$ TeV	y-axis: $d^2 N_{\text{ch}}/d\eta dp_{\perp}$ [(GeV/c) ⁻¹] x-axis: $p_{\perp}$ [GeV/c]	1.0
/CMS_2010_S8656010/d02-x01-y02	Charged hadron $p_{\perp}$ for $ \eta = 1.1$ at $\sqrt{s} = 7$ TeV	y-axis: $d^2 N_{\text{ch}}/d\eta dp_{\perp}$ [(GeV/c) ⁻¹] x-axis: $p_{\perp}$ [GeV/c]	1.0
/CMS_2010_S8656010/d02-x01-y03	Charged hadron $p_{\perp}$ for $ \eta = 1.3$ at $\sqrt{s} = 7$ TeV	y-axis: $d^2 N_{\text{ch}}/d\eta dp_{\perp}$ [(GeV/c) ⁻¹] x-axis: $p_{\perp}$ [GeV/c]	1.0

Continued on next page

Rivet path	Description	Axis	Weight
/CMS_2010_S8656010/d02-x01-y04	Charged hadron $p_{\perp}$ for $ \eta = 1.5$ at $\sqrt{s} = 7$ TeV	y-axis: $d^2 N_{\text{ch}} / d\eta dp_{\perp} [(\text{GeV}/c)^{-1}]$ x-axis: $p_{\perp} [\text{GeV}/c]$	1.0
/CMS_2010_S8656010/d03-x01-y01	Charged hadron $p_{\perp}$ for $ \eta = 1.7$ at $\sqrt{s} = 7$ TeV	y-axis: $d^2 N_{\text{ch}} / d\eta dp_{\perp} [(\text{GeV}/c)^{-1}]$ x-axis: $p_{\perp} [\text{GeV}/c]$	1.0
/CMS_2010_S8656010/d03-x01-y02	Charged hadron $p_{\perp}$ for $ \eta = 1.9$ at $\sqrt{s} = 7$ TeV	y-axis: $d^2 N_{\text{ch}} / d\eta dp_{\perp} [(\text{GeV}/c)^{-1}]$ x-axis: $p_{\perp} [\text{GeV}/c]$	1.0
/CMS_2010_S8656010/d03-x01-y03	Charged hadron $p_{\perp}$ for $ \eta = 2.1$ at $\sqrt{s} = 7$ TeV	y-axis: $d^2 N_{\text{ch}} / d\eta dp_{\perp} [(\text{GeV}/c)^{-1}]$ x-axis: $p_{\perp} [\text{GeV}/c]$	1.0
/CMS_2010_S8656010/d03-x01-y04	Charged hadron $p_{\perp}$ for $ \eta = 2.3$ at $\sqrt{s} = 7$ TeV	y-axis: $d^2 N_{\text{ch}} / d\eta dp_{\perp} [(\text{GeV}/c)^{-1}]$ x-axis: $p_{\perp} [\text{GeV}/c]$	1.0
/CMS_2010_S8656010/d04-x01-y01	Charged hadron $p_{\perp}$ for $ \eta < 2.4$ at $\sqrt{s} = 7$ TeV	y-axis: $(1/2\pi p_{\perp}) d^2 N_{\text{ch}} / d\eta dp_{\perp} [(\text{GeV}/c)^{-2}]$ x-axis: $p_{\perp} [\text{GeV}/c]$	1.0
/CMS_2010_S8656010/d05-x01-y01	Charged hadron η integrated over $p_{\perp}$ at $\sqrt{s} = 7$ TeV	y-axis: $dN_{\text{ch}} / d\eta$ x-axis: η	6.0
/CMS_2011_S8884919/d25-x01-y01	Mean $p_{\perp}$ vs charged hadron multiplicity, $ \eta < 2.4$, $\sqrt{s} = 7$ TeV	y-axis: $\langle p_{\perp} \rangle [\text{GeV}]$ x-axis: n	8.0

End of table.

Table 10. Observables and corresponding weights used in the third tuning step.

Data Availability Statement. This article has no associated data or the data will not be deposited.

Code Availability Statement. This article has no associated code or the code will not be deposited.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] B. Andersson, G. Gustafson, G. Ingelman and T. Sjöstrand, *Parton fragmentation and string dynamics*, *Phys. Rept.* **97** (1983) 31 [[INSPIRE](#)].
- [2] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 physics and manual*, *JHEP* **05** (2006) 026 [[hep-ph/0603175](#)] [[INSPIRE](#)].
- [3] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys. Codeb.* **2022** (2022) 8 [[arXiv:2203.11601](#)] [[INSPIRE](#)].

- [4] OPAL collaboration, *Measurement of the production rates of charged hadrons in e^+e^- annihilation at the Z^0* , *Z. Phys. C* **63** (1994) 181 [[INSPIRE](#)].
- [5] DELPHI collaboration, *The next round of hadronic generator tuning heavily based on identified particle data*, [hep-ex/9511011](#) [[INSPIRE](#)].
- [6] DELPHI collaboration, *Charged and identified particles in the hadronic decay of W bosons and in $e^+e^- \rightarrow q\bar{q}$ from 130 GeV to 200 GeV*, *Eur. Phys. J. C* **18** (2000) 203 [Erratum *ibid.* **25** (2002) 493] [[hep-ex/0103031](#)] [[INSPIRE](#)].
- [7] SLD collaboration, *Production of π^+ , π^- , K^+ , K^- , p and $\bar{p}$ in light (uds), c and b jets from Z^0 decays*, *Phys. Rev. D* **69** (2004) 072003 [[hep-ex/0310017](#)] [[INSPIRE](#)].
- [8] ALEPH collaboration, *Studies of QCD at e^+e^- centre-of-mass energies between 91 GeV and 209 GeV*, *Eur. Phys. J. C* **35** (2004) 457 [[INSPIRE](#)].
- [9] L3 collaboration, *Studies of hadronic event structure in e^+e^- annihilation from 30 GeV to 209 GeV with the L3 detector*, *Phys. Rept.* **399** (2004) 71 [[hep-ex/0406049](#)] [[INSPIRE](#)].
- [10] A. Buckley et al., *Systematic event generator tuning for the LHC*, *Eur. Phys. J. C* **65** (2010) 331 [[arXiv:0907.2973](#)] [[INSPIRE](#)].
- [11] P. Skands, S. Carrazza and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 tune*, *Eur. Phys. J. C* **74** (2014) 3024 [[arXiv:1404.5630](#)] [[INSPIRE](#)].
- [12] S. Amoroso et al., *Estimating QCD uncertainties in Monte Carlo event generators for gamma-ray dark matter searches*, *JCAP* **05** (2019) 007 [[arXiv:1812.07424](#)] [[INSPIRE](#)].
- [13] A. Jueid, J. Kip, R.R. de Austri and P. Skands, *Impact of QCD uncertainties on antiproton spectra from dark-matter annihilation*, *JCAP* **04** (2023) 068 [[arXiv:2202.11546](#)] [[INSPIRE](#)].
- [14] N. Korneeva, A. Karneyeu and P. Skands, *Event-generator validation with MCPLOTS and LHC@home*, *Eur. Phys. J. Plus* **139** (2024) 653 [[arXiv:2401.10621](#)] [[INSPIRE](#)].
- [15] P.Z. Skands, *Tuning Monte Carlo generators: the Perugia tunes*, *Phys. Rev. D* **82** (2010) 074018 [[arXiv:1005.3457](#)] [[INSPIRE](#)].
- [16] ATLAS collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021 (2014) [[INSPIRE](#)].
- [17] CMS collaboration, *CMS Pythia 8 colour reconnection tunes based on underlying-event data*, *Eur. Phys. J. C* **83** (2023) 587 [[arXiv:2205.02905](#)] [[INSPIRE](#)].
- [18] T. Sjöstrand and M. van Zijl, *A multiple interaction model for the event structure in hadron collisions*, *Phys. Rev. D* **36** (1987) 2019 [[INSPIRE](#)].
- [19] T. Sjöstrand and P.Z. Skands, *Transverse-momentum-ordered showers and interleaved multiple interactions*, *Eur. Phys. J. C* **39** (2005) 129 [[hep-ph/0408302](#)] [[INSPIRE](#)].
- [20] T. Sjöstrand and M. van Zijl, *Multiple parton-parton interactions in an impact parameter picture*, *Phys. Lett. B* **188** (1987) 149 [[INSPIRE](#)].
- [21] UA5 collaboration, *Scaling violations in multiplicity distributions at 200 GeV and 900 GeV*, *Phys. Lett. B* **167** (1986) 476 [[INSPIRE](#)].
- [22] S. Uhlig, I. Derado, R. Meinke and H. Preissner, *Observation of charged particle correlations between the forward and backward hemispheres in pp collisions at ISR energies*, *Nucl. Phys. B* **132** (1978) 15 [[INSPIRE](#)].
- [23] UA5 collaboration, *Charged particle correlations in $\bar{p}p$ collisions at c.m. energies of 200 GeV, 546 GeV and 900 GeV*, *Z. Phys. C* **37** (1988) 191 [[INSPIRE](#)].

- [24] UA1 collaboration, *Hadronic jet production at the CERN proton-anti-proton collider*, *Phys. Lett. B* **132** (1983) 214 [[INSPIRE](#)].
- [25] AXIAL FIELD SPECTROMETER collaboration, *Properties of jets in high E_T events produced in pp collisions at $\sqrt{s} = 63$ GeV*, *Z. Phys. C* **25** (1984) 13 [[INSPIRE](#)].
- [26] CDF collaboration, *The underlying event in large transverse momentum charged jet and Z-boson production*, *Int. J. Mod. Phys. A* **16S1A** (2001) 250 [[INSPIRE](#)].
- [27] CDF collaboration, *The underlying event in hard interactions at the Tevatron $\bar{p}p$ collider*, *Phys. Rev. D* **70** (2004) 072002 [[hep-ex/0404004](#)] [[INSPIRE](#)].
- [28] ATLAS collaboration, *Measurement of underlying event characteristics using charged particles in pp collisions at $\sqrt{s} = 900$ GeV and 7 TeV with the ATLAS detector*, *Phys. Rev. D* **83** (2011) 112001 [[arXiv:1012.0791](#)] [[INSPIRE](#)].
- [29] CMS collaboration, *First measurement of the underlying event activity at the LHC with $\sqrt{s} = 0.9$ TeV*, *Eur. Phys. J. C* **70** (2010) 555 [[arXiv:1006.2083](#)] [[INSPIRE](#)].
- [30] ALICE collaboration, *Underlying event measurements in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV with the ALICE experiment at the LHC*, *JHEP* **07** (2012) 116 [[arXiv:1112.2082](#)] [[INSPIRE](#)].
- [31] ALICE collaboration, *The ALICE experiment at the CERN LHC*, *2008 JINST* **3** S08002 [[INSPIRE](#)].
- [32] ALICE collaboration, *Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions*, *Nature Phys.* **13** (2017) 535 [[arXiv:1606.07424](#)] [[INSPIRE](#)].
- [33] ALICE collaboration, *Multiplicity dependence of light-flavor hadron production in pp collisions at $\sqrt{s} = 7$ TeV*, *Phys. Rev. C* **99** (2019) 024906 [[arXiv:1807.11321](#)] [[INSPIRE](#)].
- [34] ALICE collaboration, *Study of flavor dependence of the baryon-to-meson ratio in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. D* **108** (2023) 112003 [[arXiv:2308.04873](#)] [[INSPIRE](#)].
- [35] J. Altmann et al., *Towards the understanding of heavy quarks hadronization: from leptonic to heavy-ion collisions*, *Eur. Phys. J. C* **85** (2025) 16 [[arXiv:2405.19137](#)] [[INSPIRE](#)].
- [36] C. Bierlich, G. Gustafson, L. Lönnblad and A. Tarasov, *Effects of overlapping strings in pp collisions*, *JHEP* **03** (2015) 148 [[arXiv:1412.6259](#)] [[INSPIRE](#)].
- [37] C. Bierlich, *Rope hadronization and strange particle production*, *EPJ Web Conf.* **171** (2018) 14003 [[arXiv:1710.04464](#)] [[INSPIRE](#)].
- [38] C. Bierlich, G. Gustafson and L. Lönnblad, *A shoving model for collectivity in hadronic collisions*, [arXiv:1612.05132](#) [[INSPIRE](#)].
- [39] C. Bierlich, S. Chakraborty, G. Gustafson and L. Lönnblad, *Setting the string shoving picture in a new frame*, *JHEP* **03** (2021) 270 [[arXiv:2010.07595](#)] [[INSPIRE](#)].
- [40] C. Bierlich, *String interactions as a source of collective behaviour*, *Universe* **10** (2024) 46 [[arXiv:2401.07585](#)] [[INSPIRE](#)].
- [41] K. Werner, *Core-corona separation in ultra-relativistic heavy ion collisions*, *Phys. Rev. Lett.* **98** (2007) 152301 [[arXiv:0704.1270](#)] [[INSPIRE](#)].
- [42] T. Pierog et al., *EPOS LHC: test of collective hadronization with data measured at the CERN Large Hadron Collider*, *Phys. Rev. C* **92** (2015) 034906 [[arXiv:1306.0121](#)] [[INSPIRE](#)].
- [43] T. Pierog et al., *EPOS 3 and air showers*, *EPJ Web Conf.* **210** (2019) 02008 [[INSPIRE](#)].
- [44] N. Fischer and T. Sjöstrand, *Thermodynamical string fragmentation*, *JHEP* **01** (2017) 140 [[arXiv:1610.09818](#)] [[INSPIRE](#)].

- [45] J.R. Christiansen and P.Z. Skands, *String formation beyond leading colour*, *JHEP* **08** (2015) 003 [[arXiv:1505.01681](#)] [[INSPIRE](#)].
- [46] T. Sjöstrand and P.Z. Skands, *Baryon number violation and string topologies*, *Nucl. Phys. B* **659** (2003) 243 [[hep-ph/0212264](#)] [[INSPIRE](#)].
- [47] J. Altmann and P. Skands, *String junctions revisited*, *JHEP* **07** (2024) 238 [[arXiv:2404.12040](#)] [[INSPIRE](#)].
- [48] PARTICLE DATA GROUP collaboration, *Review of particle physics*, *Phys. Rev. D* **110** (2024) 030001 [[INSPIRE](#)].
- [49] CMS collaboration, *Strange particle production in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV*, *JHEP* **05** (2011) 064 [[arXiv:1102.4282](#)] [[INSPIRE](#)].
- [50] ALICE collaboration, *Charm production and fragmentation fractions at midrapidity in pp collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **12** (2023) 086 [[arXiv:2308.04877](#)] [[INSPIRE](#)].
- [51] P. Eden and G. Gustafson, *Baryon production in the string fragmentation picture*, *Z. Phys. C* **75** (1997) 41 [[hep-ph/9606454](#)] [[INSPIRE](#)].
- [52] B. Andersson, G. Gustafson and T. Sjöstrand, *Baryon production in jet fragmentation and Υ decay*, *Phys. Scripta* **32** (1985) 574 [[INSPIRE](#)].
- [53] G.S. Bali and K. Schilling, *Static quark-anti-quark potential: scaling behavior and finite size effects in $SU(3)$ lattice gauge theory*, *Phys. Rev. D* **46** (1992) 2636 [[INSPIRE](#)].
- [54] E. Eichten et al., *Charmonium: the model*, *Phys. Rev. D* **17** (1978) 3090 [Erratum *ibid.* **21** (1980) 313] [[INSPIRE](#)].
- [55] B. Andersson, S. Mohanty and F. Söderberg, *The Lund fragmentation process for a multigluon string according to the area law*, *Eur. Phys. J. C* **21** (2001) 631 [[hep-ph/0106185](#)] [[INSPIRE](#)].
- [56] J.S. Schwinger, *On gauge invariance and vacuum polarization*, *Phys. Rev.* **82** (1951) 664 [[INSPIRE](#)].
- [57] G.C. Nayak, *Non-perturbative quark-antiquark production from a constant chromo-electric field via the Schwinger mechanism*, *Phys. Rev. D* **72** (2005) 125010 [[hep-ph/0510052](#)] [[INSPIRE](#)].
- [58] G.S. Bali, *Casimir scaling of $SU(3)$ static potentials*, *Phys. Rev. D* **62** (2000) 114503 [[hep-lat/0006022](#)] [[INSPIRE](#)].
- [59] Z. Komargodski and S. Zhong, *Baryon junction and string interactions*, *Phys. Rev. D* **110** (2024) 056018 [[arXiv:2405.12005](#)] [[INSPIRE](#)].
- [60] L. Bernardinis, *Tuning of PYTHIA event generator with strange hadron measurements from ALICE and test on strange jet production*, Master's thesis, Università degli studi di Trieste, Trieste, Italy (2024).
- [61] T. Sjöstrand and P.Z. Skands, *Multiple interactions and the structure of beam remnants*, *JHEP* **03** (2004) 053 [[hep-ph/0402078](#)] [[INSPIRE](#)].
- [62] M. Fieg, F. Kling, H. Schulz and T. Sjöstrand, *Tuning Pythia for forward physics experiments*, *Phys. Rev. D* **109** (2024) 016010 [[arXiv:2309.08604](#)] [[INSPIRE](#)].
- [63] ALICE collaboration, *Measurement of prompt D^0 , Λ_c^+ , and $\Sigma_c^{0,++}(2455)$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. Lett.* **128** (2022) 012001 [[arXiv:2106.08278](#)] [[INSPIRE](#)].
- [64] LHCb collaboration, *Observation of a $\Lambda_b^0 - \bar{\Lambda}_b^0$ production asymmetry in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV*, *JHEP* **10** (2021) 060 [[arXiv:2107.09593](#)] [[INSPIRE](#)].

- [65] CMS collaboration, *Charged particle multiplicities in pp interactions at $\sqrt{s} = 0.9, 2.36$, and 7 TeV*, *JHEP* **01** (2011) 079 [[arXiv:1011.5531](#)] [[INSPIRE](#)].
- [66] C. Bierlich et al., *Robust independent validation of experiment and theory: Rivet version 3*, *SciPost Phys.* **8** (2020) 026 [[arXiv:1912.05451](#)] [[INSPIRE](#)].
- [67] A. Buckley et al., *Consistent, multidimensional differential histogramming and summary statistics with YODA 2*, [arXiv:2312.15070](#) [[DOI:10.21468/SciPostPhysCodeb.45](#)] [[INSPIRE](#)].
- [68] A. Jueid, J. Kip, R.R. de Austri and P. Skands, *The strong force meets the dark sector: a robust estimate of QCD uncertainties for anti-matter dark matter searches*, *JHEP* **02** (2024) 119 [[arXiv:2303.11363](#)] [[INSPIRE](#)].
- [69] A. Karneyeu, L. Mijovic, S. Prestel and P.Z. Skands, *MCPLOTS: a particle physics resource based on volunteer computing*, *Eur. Phys. J. C* **74** (2014) 2714 [[arXiv:1306.3436](#)] [[INSPIRE](#)].
- [70] ALICE collaboration, *Measurement of the cross sections of Ξ_c^0 and Ξ_c^+ baryons and of the branching-fraction ratio $BR(\Xi_c^0 \rightarrow \Xi^- e^+ \nu_e)/BR(\Xi_c^0 \rightarrow \Xi^- \pi^+)$ in pp collisions at 13 TeV*, *Phys. Rev. Lett.* **127** (2021) 272001 [Erratum *ibid.* **134** (2025) 179902] [[arXiv:2105.05187](#)] [[INSPIRE](#)].
- [71] LHCb collaboration, *Enhanced production of Λ_b^0 baryons in high-multiplicity pp collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. Lett.* **132** (2024) 081901 [[arXiv:2310.12278](#)] [[INSPIRE](#)].
- [72] ATLAS collaboration, *Dependence of the jet energy scale on the particle content of hadronic jets in the ATLAS detector simulation*, ATL-PHYS-PUB-2022-021 (2022) [[INSPIRE](#)].
- [73] ATLAS collaboration, *Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC*, *New J. Phys.* **13** (2011) 053033 [[arXiv:1012.5104](#)] [[INSPIRE](#)].
- [74] CMS collaboration, *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](#)] [[INSPIRE](#)].
- [75] LHCb collaboration, *Measurement of V^0 production ratios in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV*, *JHEP* **08** (2011) 034 [[arXiv:1107.0882](#)] [[INSPIRE](#)].