Questions from the review by Balázs Ujvári

1. **Can your PYTHIA simulations explain the azimuthal particle correlations (CMS, Fig. 1.9), elliptic flow (ALICE, Fig. 1.10) and the ratios of the yield like K^0_s… to pion (ALICE), you listed in page 12?**

PYTHIA 8 is a widely applied tool for predicting and verifying experimental results. At its core, it is a Monte Carlo event generator, however, the results of the simulations can vary greatly depending on the parametrisation of the software. This allows for testing different fragmentation and hadronisation models.

The azimuthal particle correlations were observed already in PYTHIA 6 [*J.Phys.Conf.Ser.* 589 (2015) 1, 012001], but are also reproduced in PYTHIA 8, which includes multi-parton interactions and colour reconnection [*Chin. Phys.* C 49 (2025) 044001]. Strange hadron-to-pion ratio is reproduced well by PYTHIA with colour ropes and string shoving [Phys. Rev. C 111, 044902 (2025)], but the Monash tune is unsuccessful in predicting the strangeness enhancement.

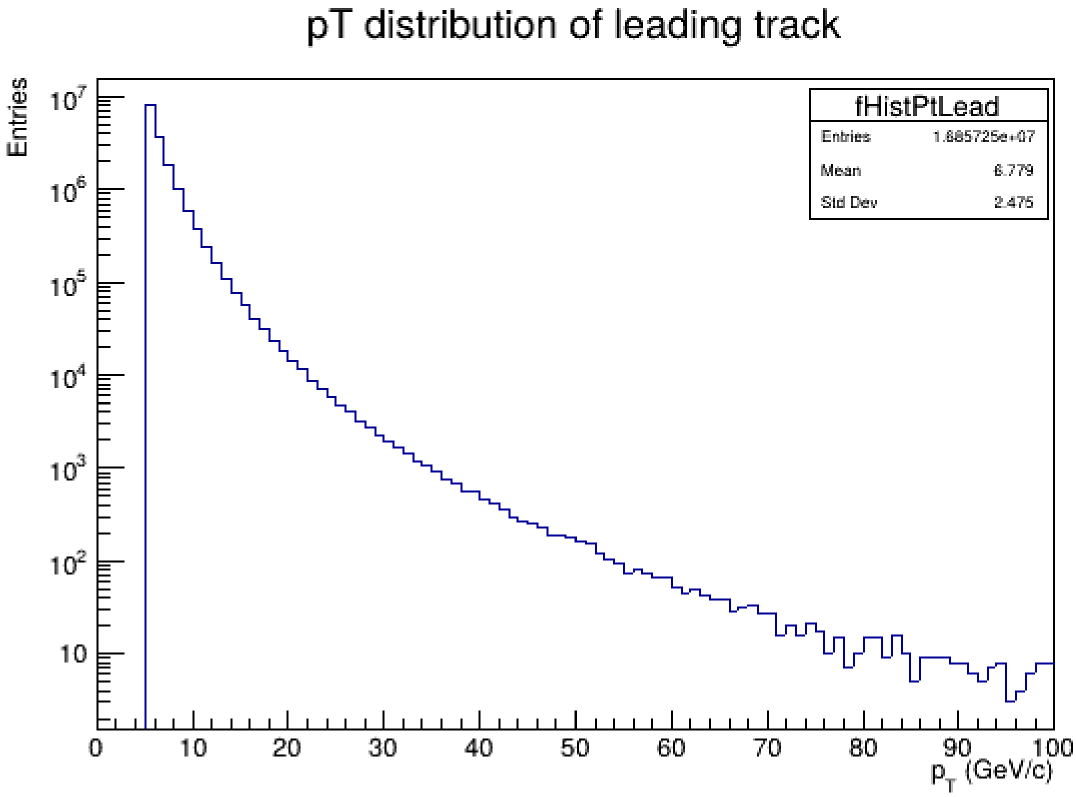
The trends of the elliptic flow effects in pp collisions are simulated correctly by PYTHIA, however, the values are not always accurate [Phys. Rev. Lett. 123, 142301 (2019)].

2. **To avoid auto-correlation effects is it the only solution (pseudorapidity and azimuth gap) or are there any statistical method without a gap to increase the statistics?**

The auto-correlation arises when constituents of a jet associated with the studied particle are also counted during the multiplicity estimation. While in principle there can be methods which overcome this (e.g. utilising MC simulations to study the extent of such effect and correct for it), the much simpler solution remains separating the measurement region and the region where multiplicity is estimated.

3. **What is the p_T distribution of the leading particle, why Fig. 2.9 stops at 40 GeV? Does ALICE have statistics till 50? For Run 3 will you expect wider range for this plot?**

The p_T distribution of the leading particle is shown in the figure below. While there are indeed events with leading particle having transverse momentum above 40 GeV, this region was not considered for two main reasons: i) the amount of data above 40 GeV is drastically lower, which would lead to high statistical uncertainties in the result; ii) The main focus of the analysis is the low-momentum region, therefore events with extremely energetic particles are not so relevant. The Run 3 data will surely increase the available data in the range of leading particle momentum above 40 GeV, which can be crucial for the analyses that rely on highly energetic particles (e.g. jet analyses, heavy-flavour spectra measurements).



4. **Page 34: "According to simulations, R_T is strongly correlated to MPI in a collision [59], therefore measuring it will indirectly classify events by MPI." In this paper we can see huge differences between MC simulations. How strongly can correlate R_T and the MPI? Are there any other papers? Can this relationship be quantified?**

The cited paper shows that the number of MPI is monotonously increasing with the value of R_T. Therefore, studying the D meson production with respect to the R_T values indirectly shows the dependence of the said production on the change of MPI (as it cannot be directly accessed in experiments). However, one drawback of the multiplicity-based event classifiers is that requiring a high charged-particle multiplicity biases the sample towards hard processes like multi-jet final states. Recently, a new observable was suggested to describe the events - flattenicity [Phys. Rev. D 107, 076012]. It is also correlated with MPI and performs well even in isotropic events (without the presence of highly energetic jets). That said, estimating flattenicity requires measuring particles in a broad range of pseudorapidity.

5. **Page 43: For charged-hadron trigger events were generated and the average N_{trans} is calculated, for jet-triggered events it was only assumed this number is same (I also think this is close, but perhaps not 7.426), why did you not calculate it in the same way? The Fig 3.3 shows small differences between the two triggers.**

For the jet-triggered events, I indeed performed the same procedure as for the hadron-triggered events, which yielded approximately the same average multiplicity. It was decided to use the same value for both sets of events afterwards. The reason behind the small differences in Fig. 3.3 is not the difference in the R_T definition, but rather that the sets of events triggered by hadrons with p_T>5 GeV and jets with p_T>10 GeV are not exactly the same.

6. **If we use different trigger threshold (for example: p_T-leading>8 GeV) will it change the Fig. 3.4?**

I studied this question at one point. Increasing the trigger threshold increases the transverse momentum value at which the transition between D-meson production in the underlying event and D-meson production in the hard process happens (shifting it close to 8 GeV in your example). The physics message of the result does not change due to this - the D mesons with p_T lower than the transition value are associated with the underlying event, while the high-p_T D mesons are still produced in the hard scattering. The universality of this interpretation is due to the measurement still happening in the multiplicity plateau domain in the transverse region. Decreasing the trigger threshold, on the other hand, would lead to a different interpretation of the result, as in that case the multiplicity in the transverse region would be dependent on the transverse momentum of the leading particle in some events (with the low-momentum leading particle). In this case, the R_T can no longer serve as a measure of the underlying event activity, as the multiplicity in the transverse region will be directly influenced by the hard processes.

7. **1% of the Run 2 MB events has p_T-leading>5 GeV particle, are there any triggers prefer at least one high-p_T track in the event?**

Utilisation of the MB trigger was a natural choice at the very beginning of the analysis. In Run 2, there was a trigger requiring high-p_T tracks in the event, however, it required *three* particles with transverse momentum of at least 3 GeV. Therefore, this trigger, while providing more events which satisfy the 5 GeV leading particle condition, would heavily distort the region with the softly produced D mesons and probably even further decrease the statistics in the first p_T bin (2-5 GeV).

8. **Table 4.4. sometimes the parameters change with p_T-D^0. How can you calculate the systematical uncertainties in this case?**

The systematic uncertainty is estimated by extracting the raw yields with all the possible combinations of the listed parameters (a total of 192 combinations). Every test is performed in each of the p_T and R_T bins *with the same set of parameters.* These sets of parameters only change from test to test, but they are the same for all invariant mass histograms within a single test.

9. **Page 66: When you calculated the acceptance-times-efficiency did you try to use 2D matrix to take into account the bin-by-bin migration in p_T too? As you did for R_T later.**

The effect of bin-by-bin migration in analyses that study individual particles is usually not taken into consideration, as it is negligible due to the accuracy of the p_T measurement of the decay products. In jet analyses, for example, detector effects and background fluctuation introduce a large uncertainty in the transverse momentum of a jet, therefore this effect cannot be disregarded and the unfolding procedure becomes necessary.

10. **Page 77, Feed-down systematic uncertainty. It was mentioned that it's a standard method as in publication 98. If I look at the 98 Table 1, and compare to your Table 4.7 (the p_T bins are different) the uncertainty from feed-down in your case seems to be much larger. Is it a bin or sqrt(s) issue or something else?**

The main reason for the different systematic uncertainties between these analyses is the presence of the 5 GeV trigger. It leads to different event selections and influences the distribution of D^0 mesons, especially in the lowest p_T bin. Because of the trigger turn-on curve, the amount of D^0 mesons in the first p_T bin is relatively low, which leads to increased statistical uncertainties of the result and also to less stability during testing against systematic uncertainty.

11. **Fig 4.17 You said at the beginning, that Monash tune does not describe well the heavy-flavour correctly, in this plot, in the 2-4 GeV interval the Monash is closer to the data than the CR-BLC Mode 2 tune, what is the reason?**

The Monash tune can adequately describe the charm mesons, while CR-BLC can also describe charm baryons.

Statistically speaking, the Monash and CR-BLC Mode 2 tunes are both consistent with the data. The current accuracy doesn't allow to clearly distinguish between the two models and more precise measurements with a larger amount of data are needed.

12. **Page 89: "This behaviour may be explained by the fact that…" Page 94: "...hinting at later formation times for baryons…" What is needed to prove this?**

These assumptions logically follow the known behaviours in thermodynamics (higher temperatures correspond to earlier stages of collisions) and, therefore, are used to interpret the observed effects. These can be verified by independent measurements; in principle, it can be estimated with hadron gas radiation measurements, but this would be a complex analysis.